

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 16:55:00 ; Search time 356 Seconds  
(without alignments)  
15156.708 Million cell updates/sec

Title: US-09-988-863A-1  
Perfect score: 2396  
Sequence: 1 gtcaccaccagcgtccgycg.....ttctcaaaaaaaaaaaaaa 2396

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2396	100.0	2396	24	AB160244
2	443	18.5	443	24	AB194086
3	425.4	17.8	728	24	AB160245
4	299.2	12.5	611	24	AB160245
5	135.2	5.6	571	21	AA69566
6	135.2	5.6	571	21	AA69566
7	65.8	2.7	96583	21	AA69566
8	65.4	2.7	300	21	AA69566
9	57.6	2.4	1356	24	AA140805

10	57.6	2.4	1356	24	ABK96801
11	57.6	2.4	1356	24	AAD31013
12	57.6	2.4	7681	24	AAD31026
13	57.6	2.4	7693	24	AAD31023
14	57.6	2.4	7693	24	AAD31023
15	57.6	2.4	8224	24	AAD31027
16	57.6	2.4	8235	24	AAD31025
17	57.6	2.4	8400	24	AAD31029
18	57.6	2.4	13917	24	AAD31037
19	57.6	2.4	14623	24	AAD31039
20	57.6	2.4	14623	24	AAD31041
21	46	1.9	4590	22	AAH24065
22	43.6	1.8	745	21	AAH24065
23	42.8	1.8	577	22	AAD02788
24	42.8	1.8	1299	22	AAD02792
25	42.8	1.8	1299	22	AAH34422
26	42.8	1.8	1763	22	AAD02791
27	41.8	1.7	164976	19	AAV21209
28	41.2	1.7	39887	22	AAK79153
29	41.2	1.7	39887	22	AAK81263
30	40.6	1.7	4590	22	AAH24065
31	40.4	1.7	116277	20	AAH20248
32	40.4	1.7	910715	20	AAH20248
33	39.2	1.6	300	23	ABV52561
34	39	1.6	3775	23	ABV52561
35	38.6	1.6	554	22	AAH10087
36	38.6	1.6	2055	22	AAH17395
37	38.6	1.6	273254	21	AAH81914
38	38.4	1.6	360	22	AAH82576
39	38.4	1.6	938	22	AAD07722
40	38	1.6	586	21	AAH08722
41	38	1.6	27681	22	AAH36497
42	38	1.6	27681	22	AAH36498
43	38	1.6	27681	22	AAH36498
44	37.8	1.6	1792	21	AAH36498
45	37.8	1.6	16633	24	ABN79985

ALIGNMENTS

RESULT 1	AB160244	standard; cDNA: 2396 BP.
ID	AB160244	
AC	AB160244	
NC	29-JUL-2002	(first entry)
DT	Arabidopsis thaliana	PMV encoding cDNA seq ID NO 1.
DE	Thale cress; PMV; phosphomevalonate kinase; plant; herbicide;	
XX	growth regulator; enzyme; gene; ss.	
KW	Arabidopsis thaliana.	
OS	Arabidopsis thaliana.	
XX	Key	Location/Qualifiers
XX	FT	685..2202
XX	FT	CD5
XX	FT	/*tag= a
XX	FT	/product= "PMV"
XX	DE10057755-A1.	
XX	23-MAY-2002.	
XX	22-NOV-2000; 2000DE-1057755.	
XX	22-NOV-2000; 2000DE-1057755.	
XX	(PARB ) BAYER AG.	
XX	Meissner R, Lechelt-Kunze C;	
XX		

S. cerevisiae pren  
Yeast phosphomeval  
Operon D DNA encod  
Operon A DNA encod  
Operon B DNA encod  
Operon C DNA encod  
Operon E DNA encod  
Operon G DNA encod  
Plastid transform  
Plastid transform  
Plastid transform  
Yeast AAD9604-asso  
Arabidopsis thalia  
S. cerevisiae ERG  
Candida albicans p  
Candida albicans p  
Methanococcus jan  
Human immune/haema  
Human immune/haema  
Yeast AAD9604-asso  
Borrelia burgdorfe  
Borrelia burgdorfe  
Human prostate exp  
Drosophila melanog  
Human cDNA clone ( Human cDNA sequenc  
Chlamydia pneumoni  
Human polynucleoti  
Human secreted pro  
Fusarium venenatum  
Human cardiovascular  
Human immune/haema  
Human secreted pro  
Human chemically m











```

Db      542 TTGCTGTGTGACGCTGC 557

RESULT 7
ID      AAF22297/c
AC      AAF22297 standard; DNA; 96583 BP.
XX
XX      AAF22297;
AC
XX      20-MAR-2001 (first entry)
DT
XX      BAC containing repeats from centromeres 1-4 #20.
DE
XX      Centromere; microsome; vector; ds.
XX
XX      Arabidopsis thaliana.
OS
XX      MO200055325-A2.
PN
XX      21-SEP-2000.
PD
XX      17-MAR-2000; 2000WO-US07392.
PE
XX      18-MAR-1999; 99US-0125219.
PR      01-APR-1999; 99US-0127409.
PR      18-MAY-1999; 99US-0134770.
PR      13-SEP-1999; 99US-0153584.
PR      17-SEP-1999; 99US-0154603.
XX
XX      (UYCH-) UNIV CHICAGO.
PA
XX      Preuss D, Copenhagen G, Keith K;
PI
XX      WPI; 2000-587529/55.
DR
XX
XX      Recombinant DNA construct comprising a plant centromere, useful for
PT      producing stably inherited microsomes which can serve as vectors for
PT      the construction of transgenic plant and animal cells
PS
XX
XX      Claim 102; Page 716-738; 1449pp; English.
XX
XX      The present invention relates to a recombinant DNA construct of a plant
CC      (Arabidopsis thaliana) centromere. The constructs are useful for
CC      producing stably inherited microsomes which can serve as vectors for
CC      the construction of transgenic plant and animal cells expressing
CC      selected proteins such as hormones, enzymes, interleukins, clotting
CC      factors, cytokines, antibodies, and growth factors.
XX
XX      Sequence 96583 BP; 29910 A; 18944 C; 18341 G; 29388 T; 0 other;
SO

Query Match      2.7%; Score 65.8; DB 21; Length 96583;
Best Local Similarity 61.3%; Pred. No. 1.5e-07;
Matches 106; Conservative 0; Mismatches 67; Indels 0; Gaps 0.

QY      100 AAAACCGTATTAACCGTTGGTTCAAGAGTGCCGGAACCGAACCAGCCCGTAACCGAATATC 159
Db      202719 AAACCCGAAGTAATGATGATGATTAACGATTAACCAACGACCAACCAACGACATGTGATTG 20220
QY      160 CTCAAAGAAATTTGCCGATGCGGTTGCTACTGCTTCAAAACCTCGGTCGCCGAGACCGAAA 219
Db      202119 CCCAAATTAATTTGGCGCTTGGTTGGTTTATCTCTTCGGTGTAAAGCCGGAAGACCGAAA 20160
QY      220 CTGTCGGTTTTTTTGGGTTGCGGTTCTCGGTTCTCTTCGGAATCCAGGCCCTA 272
Db      20159 ATTTCGGGTGATGCTGCGCGGTTGATGCGTTTAATCCGAACCTCCAGGCCCTA 20107

RESULT 8
ID      AAC37035
AC      AAC37035 standard; DNA; 300 BP.
XX
XX      AAC37035;

```

DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 15946.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131448.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149427.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.

27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152353.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159285.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.7%; Score 65.4; DB 21; Length 300;  
Best Local Similarity 98.5%; Pred. No. 7.7e-09;  
Matches 66; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2096 TTTTGCCCTTTGGTGGAGAGATCCACATGGGCTTGCCCTAGAAAGTGGTATCCAC 2155  
DB 234 TTGGGCGCTTGTGGTGGAGAGATCCACATGGGCTTGCCCTAGAAAGTGGTATCCAC 293  
QY 2156 GAACCAC 2162  
DB 294 GAACCAC 300

## RESULT 9

AA140805  
ID AA140805 standard; DNA; 1356 BP.

XX AC AA140805;  
XX DT 03-OCT-2002 (first entry)  
XX DE Nucleic acid relating to the production of prenyl alcohol SEQ ID No 29.

XX XX Prenyl alcohol; mutated cell; squalene synthase gene; geometric isomer;  
KW Industrial synthesis; isoprenoid-terpenoid compound; ds.  
XX OS Saccharomyces cerevisiae.  
XX PN WO200253747-A1.  
XX PD 11-JUL-2002.  
XX PF 20-DEC-2001; 2001WO-JP11215.  
XX PR 28-DEC-2000; 2000JP-0401701.  
PR 28-DEC-2000; 2000JP-0403067.  
PR 18-SEP-2001; 2001JP-0282978.  
XX (TOYT ) TOYOTA JIDOSHA KK.  
XX PI Ohto C, Obata S;  
XX WPI; 2002-548086/58.  
XX PT Production of prenyl alcohols by culturing translationally-active  
PT mutated cells with reduced squalene synthase gene to express less  
PT transcriptional product -  
XX PS Disclosure; Page 237-238; 262pp; Japanese.  
XX CC The invention relates to a method for producing prenyl alcohol comprising  
CC culturing mutated cells, having been mutated so as to reduce the amount  
CC of transcriptional product of squalene synthase gene transcriptional  
CC activity, and then collecting prenyl alcohol from the culture medium. The  
CC method is for the production of prenyl alcohols, which is for use in  
CC industrial synthesis of isoprenoid-terpenoid compounds particularly  
CC physiologically-active prenyl alcohol geometric isomers. This  
CC polynucleotide sequence represents a nucleic acid sequence relating to  
CC the method for producing prenyl alcohol comprising culturing mutated  
CC cells.  
XX SQ Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;

Query Match 2.4%; Score 57.6; DB 24; Length 1356;  
Best Local Similarity 50.7%; Pred. No. 3.7e-06;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1337 ATCTAGATGTTATCCATATGATGATGACACAAAGCTCATGTTGTCACAAAGGAGTGC 1396  
DB 551 ATAGAGAGTTATTCATATTTAGACAAAGTTCATGTCATGTCAGAGCTCAAAATG 610  
QY 1397 GAATGGGTTGATGTCAGCTGCTGCTATGGAAGTACAGCTTATGCTTCCTC 1456  
DB 611 GAAGCGGTTGATGATGACGGCGCAGCATATGATGATCATGATATGAAAGATTCCAC 670  
QY 1457 CAGAAGTCTTCTCATTTGCTAGGTTGCAGTATGAGTCTCCATTAAATGAGTTATTG 1516  
DB 671 CCGCATTTATCTCTAATTTGTCAGATATGGAAGTCTACTTACGGCAGTAACTGGCGC 730  
QY 1517 GTACAAATTTTGAAGGAAATGGAGCAATTAAGAACTGACTTCTTTA;CACCACATGA 1576  
DB 731 ATTGGTTGATGAGAAAGACATGGAATATTACATTAATAAGTAAACATTACTTGGGAT 790  
QY 1577 TGAATCTTTTCTTGAGAGAACCTGGAAGTGT 1608  
DB 791 TAACTTATGATGGCGCATATTAAGATGT 822

## RESULT 10

ABK96801  
ID ABK96801 standard; DNA; 1356 BP.

XX AC ABK96801;  
XX DT 24-SEP-2002 (first entry)

XX S. cerevisiae prenyl diphosphate synthase gene #20.  
 DE  
 XX  
 KW Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;  
 KW hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; gene; ss;  
 KW isopentenyl diphosphate delta-isomerase; mevalonate kinase;  
 KW mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN WO200253746-A1.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-JP11214.  
 XX  
 PR 28-DEC-2000; 2000JP-0403067.  
 XX  
 PA (TOYT) TOYOTA JIDOSHA KK.  
 XX  
 PI Ohno C, Obata S, Muramatsu M, Nishi K, Totsuka K;  
 XX  
 DR WPI: 2002-537944/57.  
 XX  
 PT Production of prenyl alcohols by culturing a transformant transferred  
 PT with e.g. prenyl diphosphate synthase gene, for use in industrial  
 PT synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds  
 XX  
 XX Example 17; Page 269-270; 335pp; Japanese.  
 XX  
 CC The invention relates to a process for producing a prenyl alcohol  
 CC comprising: (A) construction of a recombinant by transferring an  
 CC expression recombinant DNA or a DNA for genome integration into a host  
 CC which contains prenyl diphosphate synthase gene or its variant; and  
 CC (B) collecting product from the culture medium. Also described is: (1) a  
 CC method for producing prenyl alcohol in which the expression  
 CC recombinant DNA contains: (a) a hydroxymethylglutaryl-CoA reductase  
 CC gene or its variant; or (b) an isopentenyl diphosphate delta-isomerase  
 CC gene; (2) a process for producing geranylgeraniol, comprising:  
 CC (a) constructing a recombinant by transferring an expression  
 CC recombinant DNA or a DNA for genome integration into a host which  
 CC contains hydroxymethylglutaryl-CoA reductase gene or its variant; and  
 CC (b) isolating the product; (3) a process for producing farnesylgeraniol  
 CC in which the expression recombinant DNA also contains an isopentenyl  
 CC diphosphate delta-isomerase gene; mevalonate CoA acetyltransferase gene,  
 CC hydroxymethylglutaryl-CoA synthase gene, mevalonate kinase gene, or  
 CC mevalonate diphosphate decarboxylase gene. The methods are used for  
 CC the production of prenyl alcohols, particularly for use in industrial  
 CC synthesis of isoprenoid-terpenoid compounds e.g. physiologically-active  
 CC prenyl alcohols including geometric isomers. ABK96780-ABK96897  
 CC represent prenyl diphosphate synthase genes and related PCR primers used  
 CC in the methods of the invention.  
 XX  
 XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;  
 SQ  
 Query Match 2.4%; Score 57.6; DB 24; Length 1356;  
 Best Local Similarity 50.7%; Pred. No. 3.7e-06;  
 Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

DB 731 ATTTGGTTGATGAGAGACTGCAATATTACCATTAAGTAACCATTTACCTCGGAT 790  
 QY 1577 TGAATCTTTCTCTTGAGAGACTGGAAGTGT 1608  
 DB 791 TTAACCTTATGATGAGGCGCATATTATTAAGATGCT 822  
 RESULT 11  
 AAD31013  
 ID AAD31013 standard; DNA; 1356 BP.  
 XX  
 AC AAD31013;  
 XX  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE Yeast phosphomevalonate kinase (ERG8)orf.  
 XX  
 KW Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;  
 KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;  
 KW transgenic plant; yeast; phosphomevalonate kinase; ERG8; PMK; ds.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN WO200210398-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 31-JUL-2001; 2001WO-US24037.  
 XX  
 PR 31-JUL-2000; 2000US-221703P.  
 XX  
 PA (HAHN/) HAHN F M.  
 PA (KUEH/) KUEHNLE A R.  
 XX  
 PI Hahn FM, Kuehnle AR;  
 XX  
 DR WPI: 2002-217122/27.  
 XX  
 PT Use of specific genes of mevalonate and isoprenoid biosynthetic  
 PT pathways, for providing a cell with herbicide or antibiotic resistance,  
 PT and for providing transformed cells having increased isoprenoid  
 PT production  
 XX  
 PS Disclosure; Page 116-117; 193pp; English.  
 XX  
 CC The invention relates to the use of specific genes of the mevalonate  
 CC and isoprenoid biosynthetic pathways and inactive gene sites  
 CC (pseudogene). Genes of the invention are used to enhance biosynthesis  
 CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)  
 CC and isoprenoid pathway derived products in the plastids of transgenic  
 CC plants and microalgae, for producing herbicide or antibiotic resistant  
 CC transgenic plants and microalgae, for providing transformed cells with  
 CC increased isoprenoid production compared to non-transformed cells, and  
 CC for providing a cell with an inserted polynucleotide sequence encoding  
 CC one or more products of interest. The present sequence is yeast  
 CC phosphomevalonate kinase (PMK; ERG8) EC 2.7.4.2. encoding orf.  
 XX  
 XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;  
 SQ  
 Query Match 2.4%; Score 57.6; DB 24; Length 1356;  
 Best Local Similarity 50.7%; Pred. No. 3.7e-06;  
 Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Db 671 CCCGATTAATCTCTAATTTGGCAGATATTTGAGAGTGTCTACTTACGGCAGTAACCTGGCGC 730  
 Oy 1517 GTACAAATTTTGAAGGAAATGGACATATAGAGAACTGAGTCTCTTACCACACTGA 1576  
 Db 731 ATTTGGTTGATGAAGAAAGACTGGAATATTTACGATTTAAAGTAACCATTTTACCTTCGGGAT 790  
 Oy 1577 TGAATCTTTTCTTGAGAACCTGGAAGTGT 1608  
 Db 791 TAACCTTATGATGGCGGATATTAAGATGT 822  
 RESULT 12  
 AAD31026 standard; DNA: 7681 BP.  
 AAD31026:  
 31-MAY-2002 (first entry)  
 Operon D DNA encoding mevalonate pathway.  
 Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;  
 isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;  
 transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;  
 mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDP;  
 acetocetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;  
 MK; PMK; HMG-CoA reductase; ds.  
 Chimeric - Saccharomyces cerevisiae.  
 Chimeric - Arabidopsis thaliana.  
 Chimeric - Streptomyces sp.  
 WO200210398-A2.  
 07-FEB-2002.  
 31-JUL-2001; 2001WO-US24037.  
 31-JUL-2000; 2000US-221703P.  
 (HAHN/) HAHN F. M.  
 (KUEH/) KUEHNLE A. R.  
 Hahn FM, Kuehnle AR;  
 WPI; 2002-217122/27.  
 Use of specific genes of mevalonate and isoprenoid biosynthetic  
 pathways, for providing a cell with herbicide or antibiotic resistance,  
 and for providing transformed cells having increased isoprenoid  
 production  
 Claim 77; Page 141-145; 193pp; English.  
 The invention relates to the use of specific genes of the mevalonate  
 and isoprenoid biosynthetic pathways and inactive gene sites  
 (pseudogene). Genes of the invention are used to enhance biosynthesis  
 of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)  
 and isoprenoid pathway derived products in the plastids of transgenic  
 plants and microalgae, for producing herbicide or antibiotic resistant  
 transgenic plants and microalgae, for providing transformed cells with  
 increased isoprenoid production compared to non-transformed cells, and  
 for providing a cell with an inserted polynucleotide sequence encoding  
 one or more products of interest. The present sequence is operon D DNA  
 encoding the entire mevalonate pathway. This operon contains  
 S. cerevisiae orf5 encoding phosphomevalonate kinase (PMK), mevalonate  
 kinase (MWK), mevalonate diphosphate decarboxylase (MDP), acetocetyl  
 thiolase (AACT) and A. thaliana orf5 encoding 3-hydroxy-3-methylglutaryl-  
 coenzyme A (HMG-CoA) synthase (HMGs) and Streptomyces sp. CL190 orf5  
 encoding HMG-CoA reductase (HMGR).  
 Sequence 7681 BP; 2127 A; 1691 C; 1837 G; 2026 T; 0 other;

Query Match 2.4%; Score 57.6; DB 24; Length 7681;  
 Best Local Similarity 50.7%; Pred. No. 9.7e-06;  
 Matches 138; Conservative 0; Mismatches 134; Indels 0; Caps 0;  
 Oy 1337 ATCTAGATGTTATCCATATGATACACAAAGCTCTCTTGTTCGCACAAAGGAGTGC 1396  
 Db 585 ATAGAGAGTTATTCATTAATTTAGCAAGTGTCTCATTTGCAAGTCAGGTAAGATTTG 644  
 Oy 1397 GAATGGGTTTGTATGTCAGCTGTCTGTCTATGAGAACTCAGCTTATGCTTCTCTC 1456  
 Db 645 GAACGGGTTTGTATGTCAGCGCGGACACATATGATCTATCAGATATAGAAATTCAC 704  
 Oy 1457 CAGAAGTCTTGTTCATTTGCTTCAGGTTCAGTACAGGTCTGCCATTAATGAACTATTG 1516  
 Db 705 CCCGATTAATCTCTAATTTGGCAGATATTTGAGAGTGTCTACTTACGGCAGTAACCTGGCGC 764  
 Oy 1517 GTACAAATTTTGAAGGAAATGGACATATAGAGAACTGAGTCTCTTACCACACTGA 1576  
 Db 765 ATTTGGTTGATGAAGAAAGACTGGAATATTTACGATTTAAAGTAACCATTTTACCTTCGGGAT 824  
 Oy 1577 TGAATCTTTTCTTGAGAACCTGGAAGTGT 1608  
 Db 825 TAACCTTATGATGGCGGATATTAAGATGT 856  
 RESULT 13  
 AAD31023 standard; DNA: 7693 BP.  
 AAD31023:  
 31-MAY-2002 (first entry)  
 Operon A DNA encoding mevalonate pathway.  
 Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;  
 isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;  
 transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;  
 mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDP;  
 acetocetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;  
 MK; PMK; HMG-CoA reductase; ds.  
 Chimeric - Saccharomyces cerevisiae.  
 Chimeric - Arabidopsis thaliana.  
 WO200210398-A2.  
 07-FEB-2002.  
 31-JUL-2001; 2001WO-US24037.  
 31-JUL-2000; 2000US-221703P.  
 (HAHN/) HAHN F. M.  
 (KUEH/) KUEHNLE A. R.  
 Hahn FM, Kuehnle AR;  
 WPI; 2002-217122/27.  
 Use of specific genes of mevalonate and isoprenoid biosynthetic  
 pathways, for providing a cell with herbicide or antibiotic resistance,  
 and for providing transformed cells having increased isoprenoid  
 production  
 Claim 77; Page 127-131; 193pp; English.  
 The invention relates to the use of specific genes of the mevalonate  
 and isoprenoid biosynthetic pathways and inactive gene sites  
 (pseudogene). Genes of the invention are used to enhance biosynthesis  
 of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)  
 and isoprenoid pathway derived products in the plastids of transgenic

CC plants and microalgae, for producing herbicide or antibiotic resistant  
CC transgenic plants and microalgae, for providing transformed cells with  
CC increased isoprenoid production compared to non-transformed cells, and  
CC for providing a cell with an inserted polynucleotide sequence encoding  
CC one or more products of interest. The present sequence is operon A DNA  
CC encoding the entire mevalonate pathway. This operon contains  
CC 5. cerevisiae orf encoding phosphomevalonate kinase (PMK), mevalonate  
CC kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetocetyl  
CC thiolase (ACT) and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-  
CC coenzyme A (HMG-CoA) synthase (HMGs) and HMG-CoA reductase (HMGrt).  
XX  
SQ Sequence 7693 BP; 2212 A; 1526 C; 1794 G; 2161 T; 0 other;  
Query Match 2.4%; Score 57.6; DB 24; Length 7693;  
Best Local Similarity 50.7%; Pred. No. 9.8e-06;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
OY 1337 ATCTAGATGTTATCCATATGATAGACAAACGCTCATTTGTTCCACAAAGGAAGTGC 1396  
DB 4338 ATAGAGAGTATTCATATTTAGCAACAAGTTCATTTGCAAGCTCAGGTTAAATG 4397  
OY 1397 GAAGTGGTTGATGTCAGCTGCTGCTATGAGAGTCAGGTTATGTCCTCTC 1456  
DB 4398 GAAGCGGTTGATGTCAGCGGCGCATATGATCTATGATATAGAAATGCCAC 4457  
OY 1457 CAGAGCTTGTTCATTTGCTCAGTTCAGTACAGTCTGCGCATTAATGAATATG 1516  
DB 4458 CCGCATTTATCTCTAATTTGCCAGATATTGGAAGTCTACTACGCACTAAACTGGCC 4517  
OY 1517 GTACAATTTTGAAGGAAATGGACAATAGAGACGAGTCTCTTACCACCACTGA 1576  
DB 4518 ATTTGGTTGATGAGAAAGACATGGAATATTCAGATTAAGTAACATTACCTCGGAT 4577  
OY 1577 TGAATCTTTCTCTTGAGAACCTGGAAGTGT 1608  
DB 4578 TAACTTTATGATGGCGCATATTAAAGATGTT 4609  
RESULT 14  
AAD31024  
ID AAD31024 standard; DNA; 7695 BP.  
AC AAD31024;  
DT 31-MAY-2002 (first entry)  
XX  
DE Operon B DNA encoding mevalonate pathway.  
XX  
KW Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;  
KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;  
KW transgenic plant; Yeast; phosphomevalonate kinase; HMGrt; HMGs; ACT;  
KW mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;  
KW acetocetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;  
KW MVK; PMK; HMG-CoA reductase; ds.  
XX  
XX Chimeric - Saccharomyces cerevisiae.  
OS Chimeric - Arabidopsis thaliana.  
XX  
PN WO200210398-A2.  
PD 07-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US24037.  
PF  
XX 31-JUL-2000; 2000US-221703P.  
PR  
XX (HAHN/) HAHN F M.  
PA (KUEH/) KUEHNLE A R.  
XX Hahn FM, Kuehnle AR;  
PI  
XX WPI: 2002-217122/27.  
DR  
XX

PT Use of specific genes of mevalonate and isoprenoid biosynthetic  
PT pathways, for providing a cell with herbicide or antibiotic resistance,  
PT and for providing transformed cells having increased isoprenoid  
PT production  
XX  
XX Claim 77; Page 132-136; 193pp; English.  
XX  
XX The invention relates to the use of specific genes of the mevalonate  
XX and isoprenoid biosynthetic pathways and inactive gene sites  
XX (pseudogene). Genes of the invention are used to enhance biosynthesis  
XX of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)  
XX and isoprenoid pathway derived products in the plastids of transgenic  
XX plants and microalgae, for producing herbicide or antibiotic resistant  
XX transgenic plants and microalgae, for providing transformed cells with  
XX increased isoprenoid production compared to non-transformed cells, and  
XX for providing a cell with an inserted polynucleotide sequence encoding  
XX one or more products of interest. The present sequence is operon B DNA  
XX encoding the entire mevalonate pathway. This operon contains  
XX 5. cerevisiae orf encoding phosphomevalonate kinase (PMK), mevalonate  
XX kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetocetyl  
XX thiolase (ACT) and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-  
XX coenzyme A (HMG-CoA) synthase (HMGs) and HMG-CoA reductase (HMGrt).  
XX  
SQ Sequence 7695 BP; 2214 A; 1524 C; 1793 G; 2164 T; 0 other;  
Query Match 2.4%; Score 57.6; DB 24; Length 7695;  
Best Local Similarity 50.7%; Pred. No. 9.8e-06;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
OY 1337 ATCTAGATGTTATCCATATGATAGACAAACGCTCATTTGTTCCACAAAGGAAGTGC 1396  
DB 570 ATAGAGAGTATTCATATTTAGCAACAAGTTCATTTGCAAGCTCAGGTTAAATG 629  
OY 1397 GAAGTGGTTTATGTCAGCTGCTGCTATGAGAGTCAGGTTATGTCCTCTC 1456  
DB 630 GAAGCGGTTGATGTCAGCGGCGCATATGATCTATGATATAGAAATGCCAC 689  
OY 1457 CAGAGCTTGTTCATTTGCTCAGTTCAGTACAGTCTGCGCATTAATGAATATG 1516  
DB 690 CCGCATTTATCTCTAATTTGCCAGATATTGGAAGTCTACTACGCACTAAACTGGCC 749  
OY 1517 GTACAATTTTGAAGGAAATGGACAATAGAGACGAGTCTCTTACCACCACTGA 1576  
DB 750 ATTTGGTTGATGAGAAAGACATGGAATATTCAGATTAAGTAACATTACCTCGGAT 809  
OY 1577 TGAATCTTTCTCTTGAGAACCTGGAAGTGT 1608  
DB 810 TAACTTTATGATGGCGCATATTAAAGATGTT 841  
RESULT 15  
AAD31027  
ID AAD31027 standard; DNA; 8224 BP.  
AC AAD31027;  
DT 31-MAY-2002 (first entry)  
XX  
DE Operon E DNA encoding mevalonate pathway and IPP isomerase.  
XX  
XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;  
XX isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;  
XX transgenic plant; Yeast; phosphomevalonate kinase; HMGrt; HMGs; ACT;  
XX mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;  
XX acetocetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;  
XX MVK; PMK; HMG-CoA reductase; IPP isomerase; IPP; ds.  
XX  
XX Chimeric - Saccharomyces cerevisiae.  
OS Chimeric - Arabidopsis thaliana.  
OS Chimeric - Streptomyces sp.  
OS Chimeric - Rhodobacter capsulatus.  
XX  
XX WO200210398-A2.  
PN

XX 07-FEB-2002.  
PD  
XX  
XX 31-JUL-2001; 2001MO-US24037.  
PE  
XX  
XX 31-JUL-2000; 2000US-221703P.  
PR  
XX  
XX (HAHN/) HAHN F. M.  
PA (KUEH/) KUEHNLE A. R.  
XX  
PI Hahn FM, Kuehnle AR;  
XX  
XX WPI: 2002-217122/27.  
DR  
XX  
XX  
XX Use of specific genes of mevalonate and isoprenoid biosynthetic  
PT pathways, for providing a cell with herbicide or antibiotic resistance,  
PT and for providing transformed cells having increased isoprenoid  
PT production  
XX  
XX  
XX Claim 77; Page 145-149; 193pp; English.  
PS  
XX  
XX The invention relates to the use of specific genes of the mevalonate  
CC and isoprenoid biosynthetic pathways and inactive gene sites  
CC (pseudogene). Genes of the invention are used to enhance biosynthesis  
CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)  
CC and isoprenoid pathway derived products in the plastids of transgenic  
CC plants and microalgae, for producing herbicide or antibiotic resistant  
CC transgenic plants and microalgae, for providing transformed cells with  
CC increased isoprenoid production compared to non-transformed cells, and  
CC for providing a cell with an inserted polynucleotide sequence encoding  
CC one or more products of interest. The present sequence is operon B DNA  
CC encoding the entire mevalonate pathway. This operon contains  
CC S. cerevisiae orf5 encoding phosphomevalonate kinase (PMK), mevalonate  
CC kinase (WVK), mevalonate diphosphate decarboxylase (MDP), acetoacetyl  
CC thiolase (AACT) and A. thaliana orf5 encoding 3-hydroxy-3-methylglutaryl-  
CC coenzyme A (HMG-CoA) synthase (HMG5); Streptomyces sp. CL130 orf  
CC encoding HMG-CoA reductase (HMGR) and R. capsulatus DNA encoding IPP  
CC isomerase (IPPI).  
XX  
XX  
SQ Sequence 8224 BP; 2228 A; 1847 C; 2016 G; 2133 T; 0 other;  
XX  
XX  
Query Match 2.48; Score 57.6; DB 24; Length 8224;  
Best Local Similarity 50.78; Pred. No. 1e-05;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 1337 ATCTAGATGTTATCCATATGATATGACACAAACGCTCATTTGCTTGACACAGGAAAGTGC 1396  
DB 585 ATAGAGAAATATTCATATATTAGACACAAAGTTGCTCATTTGCAAGCTCAGGTAATAATTG 644  
QY 1397 GAATGGGTTGATGTCAGCTGCTGCTATGGAAGTACAGCTTATGCTTCTCTC 1456  
DB 645 GAACGGGGTTGATGTCAGCGGCGAGCATATGATATGAGATATGAAAGATTCCAC 704  
QY 1457 CAGAAGCTTGTCTCATTTGCTCAGCTTGCAGTTCAGTTCAGTTCATTAATGAGTTATTG 1516  
DB 705 CCGCATTAATCTCTAATTTGCAATATGCAAGTCTACTTACGGCAGTAACCTGGCGC 764  
QY 1517 GTACAATTTTGAAGGAAATGAGACATAAGAGAACTGAGTCTCTTACCAACACATGA 1576  
DB 765 ATTTGGTTGATGAAAGAGACTGGAATATATGATTAATAAGTAACATTACCTTCGGGAT 824  
QY 1577 TGAATCTTTTCTTGAGAGAACCTGGAAGTGT 1608  
DB 825 TAACCTTATGATGGCGCATATTAAGATGCT 856

Search completed: May 3, 2003, 18:09:32  
Job time : 597 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 16:56:55 ; Search time 4228 Seconds

(Without alignments)  
16492.508 Million cell updates/sec

Title: US-09-988-863A-1

Sequence: 1 gtcgaccacagctcgcggc.....ttctcaaaaaaaaaaaaaa 2396

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_pl.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_plo.\*  
35: em\_hcg\_rtd.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2396	100.0	2396	6	AX441242
2	841.2	35.1	1785	8	AF429385
3	759	31.7	119420	8	AC079041
4	425.4	17.8	728	6	AX441245
5	299.2	12.5	611	6	AX441244
6	227	9.5	40548	8	AB011480
7	135.2	5.6	571	6	AX441246
8	123.2	5.1	2004	6	AX461344
9	123.2	5.1	89840	8	AC004705
10	123.2	5.1	92524	8	AC005398
11	108.2	4.5	114041	8	AC005957
12	93.2	3.9	81877	8	T32B20
13	91.4	3.8	54386	8	AP002068
14	87.4	3.6	96183	8	T518
15	86	3.6	108061	8	AC009526
16	86	3.6	131692	8	AC006423
17	83.8	3.5	82214	8	AC023913
18	83.8	3.5	117585	8	AC023279
19	81	3.4	61712	8	AB017061
20	81	3.4	126253	8	AC019018
21	80.8	3.4	110565	8	AC009243
22	80.2	3.3	54252	8	AB017060
23	78.8	3.3	65839	8	T17H3
24	78.4	3.3	81414	8	AB024037
25	78.4	3.3	110157	8	AF058825
26	77	3.2	85020	8	AB018113
27	76.6	3.2	2000	6	AX461275
28	76.6	3.2	95870	8	AC007396
29	76.6	3.2	132522	8	AB062089
30	73.6	3.1	42947	8	SPAC343
31	72.6	3.0	105733	8	AC012561
32	70.8	3.0	74475	8	T7M24
33	70.8	3.0	196296	8	ATCHRIV10
34	70.2	2.9	47412	6	AX059543
35	70.2	2.9	122529	8	T1J24
36	70.2	2.9	199987	8	ATCHRIV15
37	70	2.9	92260	8	AC006194
38	69.6	2.9	82098	8	AC006136
39	69.4	2.9	112862	8	AC025781
40	65.8	2.7	42112	6	AX059497
41	65.8	2.7	96583	8	ATAC009992
42	63.6	2.7	90627	8	AP000411
43	61.6	2.6	92049	8	AC004482
44	60	2.5	115641	8	FRK4
45	59.8	2.5	81341	8	AC007213

## ALIGNMENTS

RESULT 1  
AX441242  
LOCUS AX441242 2396 bp DNA linear PAT 02-JUL-2002  
DEFINITION Sequence 1 from Patent EP1209236.  
ACCESSION AX441242  
VERSION AX441242.1 GI:21690239  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress,  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS  
TITLE  
Weissner R. and Lechelt-Kunze C.  
Phosphomevalonate kinases from plants



	Db	1981	GCTGAAGGACTTCTACTTCGTGGTGTTCCCTCGACGTGGATTTGATGCCATATTTCACA	2040
	Oy	2041	ATCACCCTTAAGGGGATATCCGGCACCAAACTGCACCAGCAGCATGGAGTTGCGACAAATGTTTTG	2100
	Db	2041	ATCACCTTAGGGGATATCCGGCACCAAACTGCACCAGCAGCATGGAGTTGCGACAAATGTTTTG	2100
	Oy	2101	GCCTTGTGGTGGAGAAGAAGATCCACATGCGCTTTGGCCCTAGAAAAGTGTCATCCACGAACC	2160
	Db	2101	GCCTTGTGGTGGAGAAGAAGATCCACATGCGCTTTGGCCCTAGAAAAGTGTCATCCACGAACC	2160
	Oy	2161	ACATGATATTACTTTCAGCGGGTTTCATCAATTCACCTTGAGTAAACAACATTTGTTTCAGTGT	2220
	Db	2161	ACATGATATTACTTTCAGCGGGTTTCATCAATTCACCTTGAGTAAACAACATTTGTTTCAGTGT	2220
	Oy	2221	CCAATTAATTAGTGTGCGTCACCAAGTTGCGTTGAGTATACTGTTTTCATATAGACTGGG	2280
	Db	2221	CCAATTAATTAGTGTGCGTCACCAAGTTGCGTTGAGTATACTGTTTTCATATAGACTGGG	2280
	Oy	2281	TGCTAATTTCTTGCTGTAGACATTTTATATCCCATTTGAAGTCTTAACTCTTGGAAA	2340
	Db	2281	TGCTAATTTCTTGCTGTAGACATTTTATATCCCATTTGAAGTCTTAACTCTTGGAAA	2340
	Oy	2341	ACTTCGCGGAAAAATATAAATGAATGATTTCACATCTTCTCAAAAAAAAAAAAAAA	2396
	Db	2341	ACTTCGCGGAAAAATATAAATGAATGATTTCACATCTTCTCAAAAAAAAAAAAAAA	2396
RESULT 2				
LOCUS	Af429385	1785 bp	mRNA	linear
DEFINITION	Hevea brasiliensis phosphomevalonate kinase mRNA, complete cds.			
ACCESSION	Af429385			
VERSION	Af429385.1 GI:16417947			
KEYWORDS				
SOURCE	ORGANISM	Hevea brasiliensis.		
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Hevea.			
REFERENCE	Hallahan, D.L. and Keiper-Hrynko, N.M.			
AUTHORS TITLE	Genes involved in the biosynthesis of isopentenyl diphosphate in unpublished			
JOURNAL	2 (bases 1 to 1785)			
REFERENCE	Keiper-Hrynko, N.M. and Hallahan, D.L.			
AUTHORS TITLE	Direct Submission			
JOURNAL	Submitted (05-OCT-2001) Central Research and Development, E. I.			
	Dubont de Nemours Co., Wilmington, DE 19800, USA			
FEATURES	location/qualifiers			
Source	1..1785			
	/organism="Hevea brasiliensis"			
	/cultivar="RIM600"			
	/db_xref="taxon:3981"			
CDS	/protein_id="AFI18926.1"			
	/cell_type="latex"			
	121..1632			
	/codon_start=1			
	/product="phosphomevalonate kinase"			
	/protein_id="AFI18926.1"			
	/db_xref="GI:16417948"			
	/translation="MAVVASPGKVKLTMGTYLIERPNAGIVLSTNARFYAIYKPIVD			
	EIKRDSSMAMNQDKVLTKSPOLARSILKLISKINALOCYSSASARNPFQAVFAYNA			
	AAATTELDKKKNLVNLKLIDGLDITLLIGTDISTRNIEICEGLPLPESIALTPFS			
	SLTNVEEANGQNCKPEVARVTGLSSAAHTVAVALHLHGLVDLSSEKEKFSDL			
	DLVIIIAQTAEICIAQGKRVSGFDVSAVSYSHRVRPSPEVLSSADAGGIPLOEVY			
	SNLKGNWDHERTFMSLPMLSELILGPPGGSSTPEMWALKKMKWSDOKSOETWR			
	KLSANSALTOFNFIYSKLAEEHMDAKCYVIDSCSTANSEKIWEIOATEPGEAEVFKAL			
	LGSNSALMOIIRVMRWKGEGEAAGVIREPSCOTDTLTMMNDGVTLAIVRAAGDFNF			
	AVTIAGDSGTNNARKWSLSNTALIVREDPKNGVLEESDPRTKETITTVAFVPHIT			
BASE COUNT	554 a 351 c 406 g 474 t			
ORIGIN				

Query Match	35.1%	Score 84.1/2	DB 8	Length 1785;
Best Local Similarity	73.2%	Pred. No. 1.3e-214;		
Matches 1109;	Conservative	0;	Mismatches 398;	Indels 9; Gaps 2
684	AATGCGTGTGTGGTCTGCTCGCTGGGAAATTTTGGATGACGTGAGGCTACCTTGTACT			743
120	AATGGCTGTAGTGTCTTGTCTCCGGGTAAAGGTGTGATGACTGGGGGTACTCATATT			179
744	CGAAGAACCCAAATGCAGGGCTTGTGTGAGTACAAATGACGCGTTTACGGATTTGTGA			803
180	GGAAAGACCCCAATGACAGGGATGTGTACTCGACCAAAATGCTCATCTTACGATTTGTGA			239
804	GGCAATCCAAAGGAAAGTCAAGCTGAAGTGGGATGGAAATGGACAGATGTCAAAT			863
240	GGCTATTATACGATGAATTAACCTGTATATGTGGGATGGGATGGACATGTGAAT			299
864	AACATCCACACAGCTCTCGAGAGAAAGCATATTAACAGTCACTGTAATCATTTGACTCT			923
300	AACATCTCCCACTAGACAGGAAAGCTTTATACAAATTTGTCACTAATAAATTTATGCT			359
924	TGACGTGTGTGCAAGTGAATCAAGAAACCCCTTTGTAGAGCATGCAATACAGTAC			983
360	TGAGTGTGTCTTCAAGATGCATCAAGGAACCCATTTGTGGAAACAAGCTGCAATTTGC			419
984	TATAGTGTGTGCTAATTTGGCAACGAGAGAG--ACAAGATATCATTCACAACTCT			1040
420	TGTAGAGGTGTCAATGACACACTTGTACAAAGATTAAGAAATGTCTTAACAGGTACT			479
1041	ATTGCAAGCTCTGATATATACAAATATTAGCTTCACATGACTTTTACTCATATGCAACCA			1100
480	CTTGCAAGGTCTGATATATACAAATATTAGTACCAATGACTCTTACTCATACGAAATGA			539
1101	GATPAGAACGCTGGGCTTCCATTGACACACAGAAATCGTGGTACCCCTTGACACCJTTCG			1160
540	GATPAGAACGATGTGGACTCCCTTTGACACAGAAATCATTTGGGTGCACTTCCTTTTTC			599
1161	ATCAATCAATTCATCAATGCTGGGAGTCAAAATGTGTATTTCCAAACCGAAGTACCAAA			1220
600	CTCAATCACCTTCAATGATGAGGAAACCAATGGACAAACTCGAACCCGTGAGTACTCTAA			659
1221	AACGTGCTAGGTCTTCTTCGACGAATGACACAGCTGTGGTGGAGCTGTGTACATTA			1280
660	AACGTGATTTGGGTTCATTCAGCAGCAATGACCACTGCTGTAGTTGCTGCTTACTTCATCA			719
1281	TCTTGAGTGTGACCTATCTGATCCATGTAAAGAAAGAAATTTGGCTTCTGTGATCT			1340
720	CTTGATTTGGTTGATCTTCAATCCCTCTGTAAAGAAAGAAATTT--TTCATGATCT			773
1341	AATATGTATCCATATGATAGCACAAACGCTCATTTCTTTGACCAAGGAAAGTCCGAAAG			1400
774	TGATTTGGTACATATATATATGACCCAACTGCCATTTATTTGACCAAGGAAAGTCCGAG			833
1401	TGGGTTGATGACAGCTGTGCTGTATATGAAAGTCAAGCTTTATGTTGCTGCTCTCCAGA			1460
834	TGGATTTGATTTAGTTCTTCGCAATTTATG6CACTCATCATACGTGCGCTTCTCTCCAGA			893
1461	AGCTTGTTCATTTGCTCAGGTTGACAGTAAACAGGCTGCACATTAATGAATTTTGGTAC			1520
894	AGTGTCTTCTCTGCTCAGAGATGCTGGAAAGAAATTCATTAACAGAAATCATTTCTTAA			953
1521	AATTTTGAAGGAAATTTGGGACATATPAGAAACTGATGTTCTTTTACCACTGATGTA			1580
954	CATCTCAAAAAGGAAATTTGGGACATGAGAGCACTATGTTTCTTCCACCACTTGTGAG			1013
1581	TCTTTTCCCTTGGAGAACCGTGAAGTGGTGGATTCCTCCACACATCAATGGTATGGTCACT			1640
1014	CTGTCTACTAGTGGAGCCAGGAACTGGAGGATCTTTCACGCCATCATGATGGGTCTCT			1073
1641	AAAGAGTGTGCAAAATGTCTGATCCAGAGAAAGCAGACAGAAACTGACACAAATTTTGCAGA			1700
1074	AAAGAAATGGCAGAAAGTCTGATACGTAGAAATTCACCAAGAAACATGGAGAAA-3TTTGCAGA			1133
1701	TGCAAATTTAGAACTGGAAACTPAGCTAAACGATCTGACCAAAATTTAGCTTAAACACCAC			1760

```

Db 1134 GGGAAATTCACACTTAAAGCAATTCATTTTAAAGCTCGAGAGAACATTTG 1193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1761 GGATGTTTATCTAGCATTAAGTCTTGTAGTGTGCTTACTGTGAAAGGGGTT 1820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1194 GGAGCGGTATTAATGTGTATGATGACAGTTCACACAAAACAGAGAGTGA 1253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1821 ACATGCTACTGACCAATCAACAGACCATTTATTAAGAACTCTTAGAGCAAGAA 1880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1254 GCAGGCAATGACCAAGCCAGGAGGAGTGTAAAGCATTAATAGATCAAAATTC 1313
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1881 TATGTGAGATTCAGATTTCTTATGCTGATGGGTAGAGGCTAGCTTCGATAGA 1940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1314 CATGCTTCAGATGCAAAATTAACATGCCAGATGGGTAGAGCTGAGTGTCCGATAGA 1373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1941 GCCGAAATTCACAACTCAACTTTTGGATTCACATGATGCTGACAGAGTCTACTTGC 2000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1374 GCCGAAATTCACAGACTGACTTTTGGATTCATGATGATGATGATGATGATGATG 1433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2001 TGTGTTCTGAGCTGAGTGTGATTTGATGATTTGCAATCACTTAGGAGATTCGG 2060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1434 TGGATGTTCTGAGCAGAGTGGTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 1493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2061 CACCAAACTGACCCAGAGCTGAGTGTGACAAATGTTTGGCTTTGTTGTAAGAGA 2120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1494 TACCAATGTGCAAAACCTTGGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTG 1553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2121 TCCACATGAGCTTGGCTAGAAAGTGTGATCCAGAACACATGATTAATCTCAGAGCT 2180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1554 CCCATAGTGTGTTTGTAGAAAGCGGAGTTCACAAAGCAAGAAATCACAACAGCTGT 1613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2181 TTCATCAATTCACCTT 2196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1614 TTTTGAGTTCATATTT 1629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
AC079041/c 119420 bp DNA linear PLN 19-JAN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC F5M6 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC079041
VERSION AC079041.4 GI:10086525
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 119420)
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
          Wu, D., Maiti, R., Ronald, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
          Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
          Arabidopsis thaliana chromosome 1 BAC F5M6 genomic sequence
          Unpublished
REFERENCE 2 (bases 1 to 119420)
AUTHORS Town, C.D. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2000) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 3 (bases 1 to 119420)
AUTHORS Town, C.D. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
          On Sep 12, 2000 this sequence version replaced gi:9945156.
COMMENT Address all correspondence to: atet@tigr.org

BAC clone F5M6 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GenScan+ (Chris Burge),

```

## FEATURES

## source

<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), GlimmerA (a variant of Glimmer, see Michaela Perlea, <http://www.tigr.org/softlib/glimmer.htm>), and GeneSplicer (Michaela Perlea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers

1. 119420

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="1"

/clone="F5M6"

misc\_feature

/note="23582 nt beyond this point were not included in the submitted sequence due to an overlap with another BAC (T12021) - the overlap is larger than 23582, but due to lack of bac end sequences, we cannot be sure where the overlap ends."

repeat\_region

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"

/rpt\_family="AT\_rich"



Dp	28358	AATGGTAGTGCAGTAAGAAAGAGTGGCAATGCTGATCCAGAGAAAGGCAAGAAACCTG	28299
Oy	1686	GCAGAAATTTGTCAGATGCAAAATTTGAACTGGAACCTAAGCTAAACGATCTGACCAATT	1745
Dp	28298	GCAGAAATTTGTCAGATGCAAAATTTGAACTGGAACCTAAGCTAAACGATCTGACCAATT	28239
Oy	1746	AGCTAAGACCCACTGGAGTGTATCTGAGAGTCATTAAGTCTGTAGTGTCTTACCTTC	1805
Dp	28238	AGCTAAGACCCACTGGAGTGTATCTGAGAGTCATTAAGTCTGTAGTGTCTTACCTTC	28179
Oy	1806	TGAAA-----	1810
Dp	28178	TGAAAAGTATATTCCTTCTGCGAGTATGATTTGTAAGATATTTATTAAGAAGGGG	28119
Oy	1811	-----AGTGGGTTCATGCTACTGAAAC	1835
Dp	28118	CAGTTTGGTTCATTTGACTTTTGTATCTCATGTTTAAAGTGGGTTCATGCTACTGAAAC	28059
Oy	1836	AATCAACGAGCATTTATTAAGAACTCTTAGAGCAAGAGAAAGTATGTTAGAGTACAG	1895
Dp	28038	AATCAACGAGCATTTATTAAGAACTCTTAGAGCAAGAGAAAGTATGTTAGAGTACAG	27999
Oy	1896	AATCTTATAGCTCAGATGGGTGAGCGGCTTAGCGTTTC-----	1934
Dp	27998	AATCTTATAGCTCAGATGGGTGAGCGGCTTAGCGTTTCGTAAGATTTGTCTGTT	27939
Oy	1935	-----	1934
Dp	27938	TTGATATGTTGCTTCAGATCTCAAGCGACCATTTGTGATGATTACTTAACTGCTGTGCAT	27879
Oy	1935	-----	1934
Dp	27878	AATATACTCGTCTGGCTGTAAATCTTAACAGCATACATGATATAGGGTTCCACAAT	27819
Oy	1935	-----G	1935
Dp	27818	TCCGATTTAGAAACAAATTAAGTCTTCATCATTAATTAAGCTCGAGGACCATGGAATTGACG	27759
Oy	1936	ATAGAGCCTGATCTCAAACTCAACTTTTGGATTTCTACAAATAGTCTGAAGAGTTCTA	1995
Dp	27758	ATAGAGCCTGATCTCAAACTCAACTTTTGGATTTCTACAAATAGTCTGAAGAGTTCTA	27699
Oy	1996	CTTGCCTGTTCTCTCGAGAGCTGGTGGATTTGATGTCATTTTGCATCACTTTAGGGAT	2055
Dp	27698	CTTGCCTGTTCTCTCGAGAGCTGGTGGATTTGATGTCATTTTGCATCACTTTAGGGAT	27639
Oy	2056	TCCGGCACCAACTGACCAGGACATGGAGTTGCGACAATGTTTGGCCTGTTGGTGAGA	2115
Dp	27638	TCCGGCACCAACTGACCAGGACATGGAGTTGCGACAATGTTTGGCCTGTTGGTGAGA	27579
Oy	2116	GAAGATCCACATAGGGGTTTGCTTAAAGTGTGATCCAGCAACACATGTATTACTTCA	2175
Dp	27578	GAAGATCCACATAGGGGTTTGCTTAAAGTGTGATCCAGCAACACATGTATTACTTCA	27519
Oy	2176	GGCGTTTCATCAATTCACCTTGAGTAAACAAATGTTTCACTGTGCCATTTATAGGTGC	2235
Dp	27518	GGCGTTTCATCAATTCACCTTGAGTAAACAAATGTTTCACTGTGTGCCATTTATAGGTGC	27439
Oy	2236	GTCAACCAAGTTCGTTGAGTATACGTGTTTGCATATAGACTTGGGTGCTAAATTTCTTGG	2295
Dp	27458	GTCAACCAAGTTCGTTGAGTATACGTGTTTGCATATAGACTTGGGTGCTAAATTTCTTGG	27399
Oy	2236	TGTAAAGCATTTTATACCATTTGTAAGGCTTTAACTCTTGAAAACTTGGGGAAAAATA	2355
Dp	27398	TGTAAAGCATTTTATACCATTTGTAAGGCTTTAACTCTTGAAAACTTGGGGAAAAATA	27339
Oy	2356	AAATTAAGTGTATTTCAAAATCTTCCA 2382	
Dp	27338	AAATTAAGTGTATTTCAAAATCTTCTCA 27312	

[illegible]

Query Match	17.8%	Score 425.4	DB 6	Length 728
Best Local Similarity	76.1%	Pred. No. 6,3e-103		
Matches	554	Conservative	0	Mismatches 162; Indels 12; Gaps 2;
QY 784	CGGTTTACGGGANTTGTGAAGCCAAATCAACGAGAAGTCTGAACGCTGAAGATTGGCGATCG	843		
Db 1	CGTTTATACGGTATGTGTTAAGCCAAATTCATGAGCTATCAAGCCGTGAAGAGCGGGCATGG	60		
QY 844	AAATGAGACAGATGTCTAAATTTAACAATCACACACAGCGCTCTGAGAGAAAGCATGTATTAAC	903		
Db 61	TCCTGGACCGATGTCTAAGACTAATCTCTCTCAGCTTTCCAGAGAAAGCATGTATTAATGG	120		
QY 904	TCACGAAATCAATTTTACTCTTCAGTCTGTGTCTGCAAGTGAATCAAGAAACCCCTTTGTA	963		
Db 121	TCCTGGAACATTTTAACACTTCAGTGTGTATCTTCAAGATCAAGAAACCCCTTTGTA	180		
QY 964	GAGCATGGGATACAGATATGCTATACCTCTGCTCATTTGGCA --- ACCGAGAAGGACAA	1020		
Db 181	GAAATVGGTATTCATCAATATACATACACCTGACATGCAACATTTTGACAAAGAAATTAAGAA	240		
QY 1021	GAATCATTCGCAACAACTCTTATTTGCAAGGCTCTGATATTAACAATATTAAGGCTCCAA	1080		
Db 241	GAGCATATAGATTAACACTCTTCAAGAGTCTGTGATATTAACATCTTTAGGTTGGCAATGAC	300		
QY 1081	TTTTATCATATCGGAGCCAGATAGAAATCGGCTGGGCTTCATTGACACCCAGATTCGCTG	1140		
Db 301	TTTTACTCATACAGAAATCAGATAGAAACACTGTCCTTCGGTTGACACCTGAACATTTG	360		
QY 1141	GGTACCCCTTGCAACGCTTGCAATCATTCATTCATTCGCGGAGTCAAAATGGTGTCAAT	1200		
Db 361	GCTATCTTACACACCTTTTACATCAATTCATTCATTCGAGAAATTCGAAATGGACCAAT	420		
QY 1201	TCCAAGCCTGGAAGTGAACAAACTGGCTTAGGTTCTTTCGAGCAATGACAAACACTGTG	1260		
Db 421	TGCAAACTGGAAGTGAACAAACTGGATTTGGTTATCTCGAGCAATGACAACTGTGTA	480		
QY 1261	GTTCGACGCTGTGATTAATTAATCTTGGAGTGTGACCTATCTGATCAATGTAAG ----	1315		
Db 481	GTTCGCTCTTACCTTCATTAATCTTGGTGTGTTAACTTTCCACCTCTTCTGCAATGCA	540		
QY 1316	----AAGGAAGTTTGGCGTTTCGATCTAGATGATCAATGATATGACAAACAGCT	1371		
Db 541	CACCAAGAAATTAAGAAATTCACAGATCTCGATATTTGCAATATGATAGCTTAAGGTGC	600		
QY 1372	CATTGCTTTCGACAGAGGAGAGGTCGAGAGTGGGTTGATGTACAGCTGTGCTGTATGGA	1431		
Db 601	CACGTATATTGGCCAAAGTAAGTGGCAGTGGCTTTGATGTACAGTTCGTGCTGTATGG	660		
QY 1432	AGTACGGTATATGTGGCTTCTCTCCAGAGCTCTTGTCAATTTGCTCAGGTGGCAGTAACA	1491		

DB	661	AGTCAGCGGTATGTCGTTTTCACCAAGAGCGCTTTCGGTGCATGAGCGTCAGTGA	720
QY	1492	GGTCGCC 1499	
DB	721	GGGATGCC 728	
RESULT 5	AX441244		
LOCUS	AX441244	611 bp	DNA
DEFINITION	Sequence 3 from Patent EP1209236.		linear
ACCESSION	AX441244		
VERSION	AX441244.1	GI:21690240	
KEYWORDS			
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
REFERENCE	1	Meissner, R. and Lechelt-Kunze, C.	
AUTHORS		Phosphomevalonate kinases from plants	
TITLE		Patent: EP 1209236-A 3 29-May-2002;	
JOURNAL		BAYER AG (DE)	
FEATURES	source	Location/Qualifiers	
		1..611	
		/organism="Medicago truncatula"	
		/db_xref="taxon:3880"	
BASE COUNT	167 a	135 c	121 g 188 t
ORIGIN			
Query Match	12.5%;	Score 299.2;	DB 6; Length 611;
Best Local Similarity	71.7%;	Pred. No. 5e-69;	
Matches 407; Conservative	0;	Mismatches 158;	Indels 3; Gaps 1;
QY	691	GTTGTGGCTTCTGCTCCTCGGAAAGTTTGTGATGACTGAGCGTACTTGTACTCGAGAG	750
DB	44	GGGGTGGCTTCTGCTCCTCGGAAAGTTTAAAGACCGGGCTACTGATTTTGAAGAGA	103
QY	751	CCAAATGCGAGGCTTGTGTGATGACAAATGCAGCGTTTACGCGATGTGTAAGCAATC	810
DB	104	CCTATGCGGAGCTTGTCTTCTTGTACTATGCTGCTGTTTATGATATGTCTCAACCAATC	163
QY	811	AACGAAAGTCAGACCTTGAAAGTTGGCGATGAAATGAGACAGATGCTAAATTAACATCA	870
DB	164	TATCTCTCAAACTCAACCTGATCTTGGGGCTTGGCTGCTCAGATGTCAGATTAACATCT	223
QY	871	CCAGAGCTCTCAGAGAAAGCATGTATAACTGTCACGTCAGATCATTTGCTTCAGTCT	930
DB	224	CCTCAACTCTCCAGAGAAAGCTTCTTATTAATTTAGCATCAAAAATCTTACCTCCAACT	283
QY	931	GTCGTGCAAGTATTCAGAAAGCCCTTTGTAGAGCATGCGATACAGATATGCTATAGCT	990
DB	284	GTTTCTCTCAAGTAAAGAACGAAACCTTTTGTGATAATAGCTGTCGAATATACCGGCT	343
QY	991	GCTGCTCATTTGGCAAC---CGAAGAGGCAAGAATCATTTGCACAAACTCTTATTGCA	1047
DB	344	GCGGCTTATGACAGAGCTGACAGATTAATAAGAGCACTTGTTCACAACTACTTTTGCA	403
QY	1048	GCTTGTGATATTAACATATTTAGCTGTCATGACTTTTACTCATTTATCGAAGACAGATAGA	1107
DB	404	GGTCTTGACATTAACATTTTGGGTTCCATGATTTTATTTATTTATAGGAATGAGATTGG	463
QY	1108	TGCGCTGGGCTTCATTTGACACACAGANTCGTGGGATACCCCTTGCAACGTTTGCAATTC	1167
DB	464	AGACAGCGGCTCTCCCTTTGACATCAAGATCATTTGGGCAACCTTCGCTTTGCTCCAT	523
QY	1168	ACATTCATATGCGGAGAGCAATGGTGTATTCATTCACCAACCTGGAATGACCAAACTGGC	1227
DB	524	TCTTTCATATGATGATGATGCTTAATGGAAGAAATTTGATAGCTGAAATTTGCCAAACTGGT	583
QY	1228	TTAGTCTTCTTGACAGCAATGACAAACAG	1255

Db	584	TTGGGCTCATCTGCAGCATGACAAACCG 6111
RESULT 6	AB011480	40548 bp DNA linear PLN 27-DEC-2000
LOCUS	AB011480	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MP17.
DEFINITION	AB011480 BA000015	
ACCESSION	AB011480.1	GI:2924730
VERSION		
SOURCE		Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MP17.
ORGANISM		Arabidopsis thaliana
REFERENCE		Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (sites)
AUTHORS		Kaneke,T., Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Miyajima,N. and Tabata,S.
TITLE		Structural analysis of Arabidopsis-thaliana chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned P1 and TAC clones
JOURNAL		DNA Res. 5 (2), 131-145 (1998)
MEDLINE		9834415
REFERENCE		2 (bases 1 to 40548)
AUTHORS		Nakamura,Y.
TITLE		Direct Submission
JOURNAL		Submitted (02-MAR-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT		Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see <a href="http://www.kazusa.or.jp/kaos/cgi-bin/ged/graph.cgi?c=MP17">http://www.kazusa.or.jp/kaos/cgi-bin/ged/graph.cgi?c=MP17</a> Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory), <a href="http://compro.ornl.gov/Grail-1.3/">http://compro.ornl.gov/Grail-1.3/</a> , GENSCAN (Chris Burge, MIT, <a href="http://CCR-081.mt.edu/GENSCAN.html">http://CCR-081.mt.edu/GENSCAN.html</a> ), DenGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <a href="http://www.cbs.dtu.dk/services/DenGene2/">http://www.cbs.dtu.dk/services/DenGene2/</a> ) and SplicePredictor (Volker Brendel, Stanford University, <a href="http://grem1n1.zozi.lastate.edu/cgi-bin/sp.c91">http://grem1n1.zozi.lastate.edu/cgi-bin/sp.c91</a> ). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MVA8 and the 3' clone is MCM23. Location/Qualifiers. 1. 40548 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="MP17" /clone_lib="Mitsui P1" complement(3187..4899) /note="contains similarity to unknown protein gb AAD25785.1 gene_id:MP17.2" /codon_start=1 /evidence="not_experimental" /protein_id="BAB11219.1" /db_xref="gi:10177887" /translation="MASLFSRSRLSSOSLILNIFILFLIAPASOTPPSGSIOITL NSFGAGSDSCSGGLASLDHRSKCYIRSGSKPGGYDYLIKIFCIGFSGVGLH LYSAMFLPYLIGDPTAASVFCPSLDSIVLKLSIPMAGVITLISGNCAPDLFSSV VSFTRSNNGDGLNSILGAFATFVSFYVGTICLIGSRDAVIDNSTRIYVFLVAL CCGLILIFGKVTIWMALCVITSLLYVGFVSVHFDRKRMSDQILRSKEDLAENG



CDS  
8366..8686  
/note="contains similarity to plastid-specific ribosomal protein 6"  
/gene\_id="MP17.3"  
/evidence-not experimental  
/codon\_start=1  
/protein\_id="BAB11220.1"  
/db\_xref="GI:10177888"  
/translation="MSVSAIFGTVITVAVASPVLRQFVKPLGNGGLGMVIECSSRP OKSTVAHHRKTPKTKQPMIDIRKRPVYAPLPPLPAEMSPFTLASDDGGAATAAGDLY SGAA"

CDS  
11914..13036,13164..13645)  
/note="gene\_id:MP17.4"  
/evidence-not experimental  
/codon\_start=1  
/product="disease resistance protein"  
/protein\_id="BAB11221.1"  
/db\_xref="GI:10177889"  
/translation="MTSSSSWKTGDETPDOVFINRGVELRKNFVSHLEKLRKG INAFIDTEMGOELSVLLERIEGSRILALIFSPRYTESKWCLELAKKERTEOEL VYPIFYKQVPTVKELKGFDEKRELVSTDKTKKEMKALQVPLPTGLVLEK SVSDEDEVINIIIRKVEKILNRSRSGSPKSCALDPORHOKRHEFMETLRIOLE EKLRSQDETPTTIGVGMGPKITTLAKLYEKMDRELRLVLRPIIDJHEASEEGLN YLATRFLQGLAKVENANISVQAHAHAYKQDLLEKTVYLIDNVSKKDOYDALGERN WIKRGSKITITTSKSLMTQSLVNDYEVPLSDKAIKHFIRYAFDGEAGAPGGQ GNPRLSKDFVHYTKGNPLALQMLGKELGKDESHWGLKALNDHNSPFGOSICM LORVSGSYKALSOKEKDALDIACFRSODENVYASLSDSDPSNLELVKFMNI YAGKQMDHDTLYMLSKELGREATADTRKGRNPHMHTTIIAVLDKKGGSNRSIPLD LSDTRKMCYFRRHAFAMBDLYRIKTYKTRHRCPOECSDIKLNPBGLIPLNVEYLH WLPFLKEVPQDPNFGNLYDLKLPYSEIRVWEDNDKAPLKWNLNHSKTLNTLAGL GAQNLQDELNLEGCTALKEHMDVMEKFLVFLNKGCTSLKSLPEIQLISLTLIS GSKRTFQVVIDSKLEALYLDGTALKELPDIDGRLOALVNLNMGCKLRPLDSDIG LKALBELILSGCSKINEPFTWGNMSRELILDLDETAIKDMPKILSVRLCLNKSKI SRPLDLNKSOLQMLHLKCYKNLTHVLPOLPNLOYLVNHGSSLSKTVKIPKIPM KHNPSFTPTNENLEOAKKEIIVYAEKCHILASALCRDESCYELTFRCTSPGC EMPNSFSDHATISWVEFELPRMNNHRLSGIALCYVYFRKNCSHANLIVKSCEDN GEGSSSITWAKVSLIEDNQEVEVDHVFIGNYCLDFIKLVGQGGPKCAPTKAS LEFSVATGAGATLEVLKSGFVFEPEPENNRPSPRNDVAGKVINTPPSANGCFK DQAKNSGKGMQWYTIENSNTINISEAHSSQKTGFGSTLLC"

CDS  
20792..21898)  
/note="gene\_id:MP17.5"  
/codon\_start=1  
/evidence-not experimental  
/product="disease resistance protein-like"  
/protein\_id="BAB11222.1"  
/db\_xref="GI:10177890"  
/translation="MEPPARVTPSIKADCSHVNICEETVLSHLSLAALREG ISVPDACLQETKFTKQNPPLTDGARVLYVSDVEYFEDPMPFKLKYIQGON NCHVAVPEYGVDSLTRYGMANSWLEAKTSHOSKFEVYIILFSFKKILSNVLT DSELVEIVRVDYGRKYPAERGVIAARLEIKLYKQHRDRIISGIMWPKIKGTLT AKAVNHSMTDADASCFTIENPEAFKEGLHRLKERIKLDEFDIESSYIMRPTL HBDKLYDKRILVYLDVDRSLIAAESFLKRLDFWFGSGLIITISVKOYAPARONIT YVOGLNVHEALQISOSVGINPEQNDKLSMKYLDVYNGPMLAISITVIGRELMKS EMEPAFELKRCPRKIDQVILKNAYSALSNDKENYVLDIAFFKGTAVYVQMLEES HIFPRALIDVLVDCVLTISENTVOANNLIDQTCOIFNGELETCTRMERPIRIRLL EYDELGESETKAMPKSGVLAHIESIFLDTSNVKEVDHAFKNMFLKPLKINSC SKYISGLNPKGLDSLPEYELRLHWNENYLPQDFDGHVYKLSMPSQLKIGTR VYDLVAKRLILSHSLOLVECDILYAOINIELIDOGCTGRPPTSGLOMLRYVNL SGCTEIKGSGYVPNIIEELHOGTRIRLEIPINATHPPVKRPPDKKIMMLNFSVE HIDECVYNLATVTSNNHYMGKLVCLNMKTCGNLKLPMVLSBKAVLYLSCSLE KLMGFPRNLKRLYVGATAIRELPQNSLEFINAGCKLKLINLDFEQLPHFIFSN KYRFSOVYIAEVEKGLVASLAKAROEELIKAPVEICIPMPTORSTROAGRAN TDLVPMQKPIGSGSVYVFDODYHNHVGIRIKVGMKWTMNOQDITVIRFCQW APTEAPKVYADHIFVLYDTKMHPSDESNHISIMAHVEVFEETHTVSGENPLGASCKV TECGVEVITAAATGDTSVSGITRESEITITIEKEDTIIIDEDTPLLKRKEPENTRSS

CDS  
22878..22949,23030..23148,23223..23281,23363..23709)  
/note="emb|CAB16816.1|  
/gene\_id="MP17.6"  
similar to unknown protein"  
/codon\_start=1  
/evidence-not experimental  
/protein\_id="BAB11223.1"  
/db\_xref="GI:10177891"  
/translation="MNEISDLGMNHNKPCYKELRHNPCYCEKKIPRTAEGLYNHP FMWEYCPSHDDGTPKCCSCERELGTYVWMLADPFRNLCEBMDSATMDSDECPL HFEIRREFEGHLMKIEEPVYLVENKALNKAKEKKIDKQDQCLMAYRGICLSEQ IYTSQGVRRMLNKOILDTVESQVYKRCVTVALILYGLPRLTYGILAHENMHA YLRNGLRYMLNVLEGLCOVGYMVECOYVFPATATASSSSRTPLSTTSKY DPSDEPKRLVNCCKHDIETDESFPDGGFRKKNKMAANNHSLKDTLKILISKRPQ YSKL"

CDS  
25693..26888,27128..27239)  
/note="contains similarity to microfilin-like associated protein 1"  
/gene\_id="MP17.7"  
/codon\_start=1  
/evidence-not experimental  
/protein\_id="BAB11224.1"  
/db\_xref="GI:10177892"  
/translation="MSVTVAGVSAVAAREKLGIGQTVRRVYMPKAPMAEAE DDVVRQKQVSLDRAFPKNDQGVSAKDDPRLAKTVENRDEVRADARRRQAEI IYEESEENSDNEDDDEDDDELRERRIRKRLNRAOEALPLLEEDDEIOEEB EEESEYETJDSDDMGIMIRPYVYKAPRTDIARELHLEBEELAREKLKLEOR KLEQTVIVEVRKDEIRKNLLEBANIGDVEITAELEAEVEWKEIRIGRLIRE RDAREMLAREEIEKRLNMTDQERDMKRNKPKPLSAQPKRMNPKRYKRGAFQ ADPDEAGSAGDGIPOPDFSAPTGEDRLDKSILPVMGVKHFGRSGRTKWHLVNED TQWMSNPTSNPLREKYNKKAAGAPAKPKSKMKMDWES"

CDS  
28445..31403,31762..32841)  
/note="gb|AAC35233.1|  
/gene\_id="MP17.8"  
similar to unknown protein"  
/codon\_start=1  
/evidence-not experimental  
/protein\_id="BAB11225.1"  
/db\_xref="GI:10177893"  
/translation="MTSDSEFRVQIRLPMITRTYKICNHPPLDGFVFLYTLH RCVPLFLAPLAVASPVLYCTVFLGTLISFGSPNLPLEIKDEIPIEAPLPTESRD AAVTVVDRGDSFTVEVSFGAEKVLLEDGNDADERYLDSQSEVEDDGRPDYRPLV DETLDEIKDTHVREFEKAFILIDVEKGGREBKILENGTAQOSRTGSLYERMD OMVSPSPMRHREDEDDADRDSDLSQDGAESSSPASMTDIIPLMDELPLI LLSAPTRCTVDCGSDASBGRPHRSSPQEMSPQSGSHGEBDNEDEDEDEE EDEBEKEREKEDDESATIKWTEADQNRVMDLSLELRNORLENLARRARRNM RLMAERNLIDFDSADIPFNMPPISTARHNPVSYDSYDMPIZGSAPSIMRARNPF DLPEPNEKRPDLKGDGFQEEFSQPKPMPRHRHFSFGVSMGLGKPHDLRPEFV LRLANEGTSYRPFEROLSEVESKYSIPDTESYCTVLEDEKDEKDNNAARETKIA KYDWSNDNEENNHSASDHDEENSHASDDEKSHSSSDSPDEVDADSKLHHVAAE IYVSGETHHROSDMMEGTSDKGLDEYSDGSSISEKEKIRDISDEPAULISQV YDLHEIAGASSLPREBELTINMARGVEDYHNDKARAESFTIAPHSDLSAHLVLCG LDDGDHEEYVYSSPPSGSRPFSVSVDYLPDLKEKGELEEBEWEKERYVSESI GDELDHSTENETRTSSEVGSNHYTASGVASVLMRHSPLPEESPLVYADLETYNK SVVEELIYEEEAOKOKDEVSQMDTIFADIPDISYASLSGAEVYETHFNFNEDVAOL EDEPVSLHVADEETHNDQMTIDEVSNASANOVSSEETSPESDERELTIWSDSVV

Query Match  
Best Local Similarity 95.9%; Pred. No. 1,1e-49;  
Matches 233; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

18 GGGCGACCTTCTCTCTTAAAGACACATATATAGAACAACTGGGAGAGA 77  
|||||  
24896 GGGCGACCTTCTCTCTTAAAGACATACATATAGACAAAGCAACTGGGAGAGA 24955  
|||||

78 TGAAGATGAGATGGTGAAGAACAACCGTATACCGTTGGTTCAAGAGTGGCGAACC 137  
|||||

Db 24956 TGAAGATGAGATGGTGAAGAACAACCGAATACCGTTGGTTCAAGAGTGGCGAACC 25015  
|||||











REFERENCE 1 (bases 1 to 114041)  
 AUTHORS Rounsley, S.D., Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Mason, T.M.,  
 Shen, M., Rinning, C.M., Fraser, C.M., Somerville, C.R. and Venter, J.C.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 114041)  
 AUTHORS Lin, X.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 REFERENCE 3 (bases 1 to 114041)  
 AUTHORS Town, C.D. and Kaul, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org  
 COMMENT On Apr 18, 2002 this sequence version replaced gi:6598502.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..114041  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="2"  
 /map="m1398"  
 /clone="T15J14"  
 1..10045  
 /note="Overlap with BAC clone T26120  
 (AC005396:75170..85214)."  
 misc\_feature  
 complement(60..82)  
 /rpt\_family="(TAA)n"  
 repeat\_region complement(1075..1102)  
 /rpt\_family="AT\_rich"  
 repeat\_region 1417..1446  
 /rpt\_family="AT\_rich"  
 repeat\_region 1481..1502  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(1836..1867)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(3450..7066)  
 /gene="At2g14970"  
 /note="T15J14.1; related to En Spm transposon family of  
 maize"  
 mRNA  
 complement(join(3450..3728,3810..4172,4473..4769,  
 4847..4885,4927..5217,5452..5532,5790..5993,6080..6673,  
 6899..>7066))  
 /gene="At2g14970"  
 complement(join(3450..3728,3810..4172,4473..4769,  
 4847..4885,4927..5217,5452..5532,5790..5993,6080..6673,  
 6899..7066))  
 /gene="At2g14970"  
 /codon\_start=1  
 /product="En/Spm-like transposon protein"  
 /protein\_id="AAD03355.1"  
 /db\_xref="GI:4115353"  
 /translation="MASQAKOVFFLRDSSSWIVANGSPRRREKENDVTLIDICP  
 LPTVVDIDLEKRGKRGKRAATKCNDEYVGTLEPAVQTVVPIVEED  
 IDGETEKEADINGETEKEADINGETEKEADINGETEKEADINGETEKEADIN  
 AEEPEGELEISGDEDDVOPKRGKPTRMKDIAKPAVRVEYTMGEPISGKSVK  
 LASYAGALVREHVPITINRMKIGBEIRITLTKSVQKFEIDOKAVAKIKONGCLM  
 RSKSRQVTRERAKTNOQRWNLPRKVNPREMKPYKSKSPERKTKGSDPEYVR  
 LKVVYKSKTKYGTGPKKAEIVSSGPOSGNENODSISQILGPDNPSRLRANGNN  
 KTKLACFOVSKCAEMQKODLOQKAVNELQEVIDIKTNVNTCLLHKRANOSROSA  
 EYGETSARVNSQKOPKILMDWAGDAIVVEGCIISDDEIVNGSRIGPTVKYL  
 IDTAIVEAYLMPALINMEIKAVGMIAPVAMCVSLERKLNPEIDIAOSFVAMP  
 KVSVENTDSAVOKAPLOSOSYPTTSASKGSAASQONGATEKSIAGEOONKAT  
 GKSGHNSIOVSAKSGSEAKTISLGEOPVYVPRKSHYKKTOLASOTLARSPROK  
 TSEGLKANOCKLMDISGKRVYAGRGYHSIDPOKVIHVNLGEVNAWVAVVAVDD  
 TAVVRPSDEIEYMDSSLSGSIAMPKDLVY"  
 repeat\_region  
 6730..6824  
 /rpt\_family="(CAA)n"  
 complement(7749..10105)  
 /gene="At2g14980"  
 /note="T15J14.2; related to En Spm transposon family of  
 maize"  
 complement(join(<7749..8219,8355..8663,8888..>10105))  
 mRNA

CDS  
 /gene="At2g14980"  
 complement(join(7749..8219,8355..8663,8888..10105))  
 /gene="At2g14980"  
 /codon\_start=1  
 /product="En/Spm-like transposon protein"  
 /protein\_id="AAD03356.1"  
 /db\_xref="GI:4115354"  
 /translation="MVOLGEIAERDDKKEDELLAKLYDAETPLPSCVNNHSLAIS  
 LFRLTKNGMSDKSFNDLETLPELPEPDVNLPHSLVEVKKFKLSFDPKGYOKIACV  
 DCCLEPRKKYKLCNLCNLSRMTNNHTGCVKGVQKYLIRPSIVPRKLRMRSEE  
 MARDLRHFNKSTDGKLCHPYDVTYWDONAKTPLRASERNRLGISTGFPNPK  
 KNTKYSCPVLLVNYLPPDKMKRENIMTLILPGSQQPNSIDVYLEPLIDLCIL  
 WDNEWTYDASKSFYKMLMTISDPAYGNLAGKVKVAKACILGKHTESMML  
 KFSRHVYGMHMGILPSHREPRNKSMFPGKAEHRKSRILISGLEISHNKPONNEG  
 NFQSTRRKQPESDLPVRNLDVYHVRNVASIVSTVMHSGSKSGVNRKDLVYL  
 GIRDLDHOTGKRTYLPALMSLSKSKKIFCRRLDPKPDYCSDIAGVLEDC  
 KRVIDMEOLEIEAEIVETLCLFSFPPSPFDIMVHLTVLHGREALLGVPHRRMY  
 PFERMYKVLKDPENRTARPEGCLAECLAECLQPCSEFLKRTYKVOEKADRNTEYN  
 NSILGGRPISAGTFISLTEMERKTAHLAVIONMAVPEPYV"  
 10482..10507  
 /rpt\_family="AT\_rich"  
 complement(13175..13216)  
 /rpt\_family="AT\_rich"  
 repeat\_region 13887..13922  
 /rpt\_family="AT\_rich"  
 repeat\_region 14418..14533  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 complement(14535..14596)  
 /rpt\_family="AT\_rich"  
 repeat\_region 14726..14753  
 /rpt\_family="AT\_rich"  
 complement(14941..14984)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(15051..15080)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(15282..15304)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(15639..15677)  
 /rpt\_family="AT\_rich"  
 repeat\_region 15760..15780  
 /rpt\_family="AT\_rich"  
 repeat\_region 16882..16912  
 /rpt\_family="AT\_rich"  
 repeat\_region 16934..16962  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(18558..18615)  
 /rpt\_family="(A)n"  
 repeat\_region 18688..18719  
 /rpt\_family="AT\_rich"  
 repeat\_region 18849..18895  
 /rpt\_family="AT\_rich"  
 repeat\_region 18964..18993  
 /rpt\_family="AT\_rich"  
 repeat\_region 19026..19046  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(19873..19908)  
 /rpt\_family="AT\_rich"  
 gene 20381..25506  
 /gene="At2g14990"  
 /note="T15J14.3; pseudogene, reverse transcriptase"  
 /pseudoc  
 complement(20898..20947)  
 /rpt\_family="(CAT)n"  
 complement(25525..25554)  
 /rpt\_family="AT\_rich"  
 complement(25946..25972)  
 /rpt\_family="AT\_rich"  
 complement(26681..27683)  
 /gene="At2g15000"  
 /note="T15J14.4; supported by full length cDNA:  
 complement(join(<7749..8219,8355..8663,8888..>10105))



GSNPAHLIISPEVLYDVLIGMDWLDHYRVLDCHGKVSFEFEREGRLYGVGPRTSG  
 SLVISAQAEKMIKGECEAVYLMVIMSPESVGOELPPSRSDPFTIELEBGTAPILSKAPY  
 RMVPEVMAELKLOLEDILGKGFITLNTSPMPTSYLFPKKGDSFRICDYEELRLAVY  
 KKKYPLPRIDELDQLGATCFESKIDLTSGTHQPIAEADYKTAFFRTGRHFEEVVA  
 PFGILNAPAAFMRLMNSVDFEELDEFVITFIDDLIVKSSPEEDVHLRAMEKLRRE  
 KLFLKLSGCSFMORRMGLGHIVSEGVSDPEKTEIARMDPRPTNATEIRISFGLAG  
 YRRFVGFASMAQPMTKLTKGDVPEVSPCEGEPVSLTSTVLALEPGEPEVNY  
 TDASVGLGCVLMORGVIAVASROLRKHGEGNPTPHOLEMAAVFTDHRSLKLYFTPOE  
 LNLBRMMELVADYDLIAVHPKSNMVMALSCKLNGAARGOSVEALYSEIALRL  
 CAVAREPLGLEAVRADULTVRVLAQEDBELLASRAEGSEIOPAAAGTILVIGRVC  
 VPKEELRETLISAHASMSIHGATKMTDILRYQWGMKRDVAVMAVECCOL  
 VKAEHQVPMGMSLOSLIPDMKMFITMDVVGGLVSRTKDAIWIIVDLTASARFLAI  
 RKTDAVTLAKKIVSEYIKLHGVPVSYSDDFEFTSFNRAFOAEMDTKQASFLAY  
 POTGOSEERTIOTLEDMLRMCVLDMGHMAHLSLVYVREAMDVTLWLDGSERSY  
 GADYVOETTERIRYKLNMKEADDRHSYADKREDFEYDGVLYLKMAELRDPNS  
 ISEIKSPRYIGPFRIYERGVPAVRIELPDMRAFRKVPVMSLRKCLHDEYDLAK  
 IPEDQPMTEARLVLERIKELRRKIPILIKVIMDCDGVTEETWEPRARKARF  
 KMFKEQVAA  
 18637. .21635  
 /gene="T32B20.g"  
 join(18637. .20487, 20649. .21101, 21186. .21635)  
 /gene="T32B20.g"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.g"  
 /protein\_id="AAF67364.1"  
 /db\_xref="GI:7682783"  
 /translation="MATSKDLPSPSEDLYEELRLNQSFKLMMRYLIAAESPFK  
 RFIYERALKALPGSYKLVAYLEREDLYANLTHQYDSLNNTFERGLVTHMKRP  
 RIMVYIQLTLVQOLITRTFRFALCALPVTQHDRIWEYLVFVSGNIPLETSLR  
 YRYRYLWDPSEHIEFELFVKSERQESARLASVANDKFSYIKGKTKMLWELC  
 ELLVHANVVISGLVNDVAILRIGIKRFTDEVDLADYIRKLNLEKADIDYEGM  
 MKVTVDFESYIFDVYSRFESEYAKKEMSSDEEDNEENGVEDEDEVEHNFNL  
 SVRELKRIINGELFNDNDVDLRLALEELMRRPALANSVILLRONPHNEQHRRY  
 KIFEGNAKQILTYTEAVRYVDPKAVGKPTLVAVAKLTENKDLNTRVITDKAY  
 QVNTKIVDLASVWCMAEMELRKNKGALIELMRATVAVTEVRRVAAADGMEPE  
 MKLRSLRMSFYDLDESGLTSTRAVVEKILDLRLATPOIIMNVAFLDEKPEY  
 DAFVYERGVYIFKYPVKIDIMWYILTRFYRGKTEILARLELEHVASPADAV  
 TLYQVAKLEEDYGLAKRMKVVEEATKVPQEGOKLEWLEYISRAAIFGVPTREI  
 YEQIESGLPHKQVIMCIRKPALESIGEDRARALYKSSQADPRSDPEFNKHH  
 EFEVQHNEDTYRMLKIKRVSASYSOTHTLPENMQKCKLLDVEAPKGLRAGL  
 PEDMAALERQLSTTPTPEPAKDGGRVGSAGVISQSGENGRKYTGAGGEDIELP  
 DESDESDGDHVEISQKEVPAAVFGGLARKRDEDEGEAGAAOKIGALERIKRQK  
 LMQ"  
 complement(22005. .23540)  
 /gene="T32B20.e"  
 complement(join(22005. .22139, 23232. .23540))  
 /gene="T32B20.e"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.e"  
 /protein\_id="AAF67362.1"  
 /db\_xref="GI:7682781"  
 /translation="MATSVALTSSPPVSLPLSSSSNSFFSNCFVTYTRNTRSLVA  
 IGRIRQEPTRKPLTNALFGLVAPLAVAGVALLFGPKLDELIGSTCKTVKSFQ  
 QAAKEFESLTPTEPESVASSQVATSNKEEKTEVSSSKENV"  
 25373. .26596  
 /gene="T32B20.h"  
 join(25373. .25663, 26010. .26204, 26264. .26596)  
 /gene="T32B20.h"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.h"  
 /protein\_id="AAF67365.1"  
 /db\_xref="GI:7682784"  
 /translation="MKNITFGIKIIVDGKIKNEKDCGEREKVDRKIDILLFLCKYI  
 VAKSDGNDIDRFLOKIVDVHACETFPWGFETDGMGEGIKVIMNNKCGAKISSTI  
 VEYVDVDEKIGFIDLVDVSWTRTLIKKKKLMMDMDHVDVRSDDOMKQOHLSENY  
 PHMLREHOKRLITIERODGENDVPFOEKGVVDVAVDEVDGENVPPREENVGNVPP  
 YEEVAVNEKIDGCVPPPEENVVDMTMEGVETGVEFKQVETDVGKCSNMH"  
 complement(30562. .31125)  
 /gene="T32B20.d"  
 complement(30562. .31125)  
 /gene="T32B20.d"  
 complement(30562. .31125)

/gene="T32B20.d"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.d"  
 /protein\_id="AAF67361.1"  
 /db\_xref="GI:7682780"  
 /translation="MVVNAKICIRPALHVEVNNKSEERVEQISRVGESSVGNVEELN  
 AFVDNGVAVTIFGGEIYVKGAVDEPIEETNCKVEEAMENADRIPEANSVEEAM  
 NAVIDNRHVSRRDSEEDNVDGRIRIETELSCAGERHSSQKRYVLBNMGNGGLKY  
 IKNLPLRRQCKIMLTKQISTVLRPOL"  
 complement(34511. .36010)  
 /gene="T32B20.c"  
 complement(join(34511. .34783, 34851. .34967, 35059. .35134,  
 35199. .35310, 35414. .35504, 35744. .36010))  
 /gene="T32B20.c"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.c"  
 /protein\_id="AAF67360.1"  
 /db\_xref="GI:7682779"  
 /translation="MEKVFSDSEITSGNHMSVNGMTSLNBSASEMAFNFIOESSAA  
 DGESTRACGVSVSSPPVVPVDSERYRAFLSKLNLACAAVAMKRGFTIKFQDTSGS  
 DNGANESQASLASKATPMSSAITSGLSGDEEDADETNANPTNVRVRMLIS  
 NRESARRRRKQAHLSLEIQLRVENSKLMKGLDYQTFENDASVENRYLKANIELL  
 RAKYMAEIVYRLTGFENPMFHNMPQIYSTVLSLSETSNSPDITTSQVTPTEIISGN  
 KGLIGCKMNRFTASMRVESLEHLQKIRSVGQ"  
 46929. .49951  
 /gene="T32B20.i"  
 join(46929. .47305, 47362. .48144, 48829. .49339, 49448. .49951)  
 /gene="T32B20.i"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein T32B20.i"  
 /protein\_id="AAF67366.1"  
 /db\_xref="GI:7682785"  
 /translation="MKMKDTEKIDDFAGRLSEVSTRKSASLGNTIEMPKLVKPLNSLP  
 RKYTHILALAEQVLDVNTLFEFEDIVGLKAYEETCCEDDQGLMILTKLIINQVY  
 GEVNDPGEDEEVMMEDVIDLKLKQETQESNGDPTQOADALIMHELYINENK  
 QKALKEENSDKDMWYLDNGASNMTGLKLEFESKIDKVTGKVRGDSRIDIGKGS  
 ITPFSKNDKSLLEDVYITIPOLRSIIISLGCATPAGCVRARKITLTFDEBELLIN  
 ANRNRNRYLYKYLEVKQKQCLQSLSSSSMMHMDHIGDNMKTIVNKLVLTRIKP  
 LILVQKQCTSCILKOVQKQSPQSTYSASNPDLVHVDLQGLTPPTTARKSGVPG  
 NKGIJDESDEAEKDEEDKNDHEDIODSALDGNVYVPIQEEIIPMLRSRQTRPT  
 YLKGIIILSEVEERLRLAIDEEPMDFEAEALKEMRACREESIKERNMNLVNR  
 PPOAKATGLKAVFKLRNSDGTINKYKRYAKVQYKGVGYDEVFALFEVYVQPE  
 GFYBGRKEDQYKTLNKLALYGLKQAPRANNNLNLQIMELKZYKSKSEPSLYRKGNEB  
 LLIYAVYVDDLVTRSSLRMLLEFRKEVSTFEPSDGLRLTYIGTEVLOHENGITLS  
 QERYVNLLEETKDKGCNAVRIIPMCANLKLKARKEASIDER"  
 complement(56115. .56312)  
 /gene="T32B20.b"  
 complement(56115. .56312)  
 /gene="T32B20.b"

Query Match 3.9%; Score 93.2; DB 8; Length 81877;  
 Best Local Similarity 70.2%; Pred. No. 9.8e-14;  
 Matches 125; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY	100	AAAACCGTATACCGTTCGCTCAGAGTGCAGAACCGACCGACCGTAAACCGAATC	159
DB	12513	AAACGGAATTAACCGATCGGTAGTGTGCGAACCGACCGACCGTAAACCGAATC	12454
QY	160	CTCAAGAAATTCGCGATCGGTTCGCTACTTTCAAAACCTCGGTGCGAGAACGGA	219
DB	12453	TCAATTAAGTTCGCGACCGGTTTACTCTTTCTTAATATGCACCGCGATTAACGGAAT	12394
QY	220	CTGTCGGTTTTTTCGTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG	277
DB	12393	GTTCGCTTCGTTTCGTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG	12336

RESULT 13  
 AF002068/c  
 LOCUS AP002068 54386 bp DNA linear PLN 27-DEC-2000

DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone: T803.
ACCESSION	AF002068 BA000014
VERSION	AF002068.1 GI:8051676
KEYWORDS	
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:TAMU BAC clone:T803.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	1 (sites)
TITLE	Nakamura,Y.
JOURNAL	Structural Analysis of Arabidopsis thaliana Chromosome 3. III
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 54386)
TITLE	Nakamura,Y.
JOURNAL	Direct Submission
COMMENT	Submitted (16-MAY-2000) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=T803 Genes with similarity to proteins in the databases are described in product or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SlicePredictor (Volker Brendel, Stanford University, http://gremmlin.zool.lastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is F814 and the 3' clone is F1M23.
FEATURES	Location/Qualifiers
source	1..54386
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/chromosome="3"
	/clone="T803"
	/clone_lib="TAMU BAC"
	/complement(join(827..1073,1360..1604))
	/note="gb AF06086.1
	gene_id:T803.1
	similar to unknown protein"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB01980.1"
	/db_xref="GI:9294129"
	/translation="MTRPPIVSGRSPFGVNGTSTSTSTGTGLQPLVLPQISNAQAVI EANDQANIEANDREQAVIRIVLHPLRRNGAKWEMDASIEHLVYNFDHLAA TRLKGMVSLAKSGKQSEWILSEHWRVMDYWMTPKAKAKSEKARSRLPSRDGLAH CHR"
	/complement(join(2125..2826,2909..2991,3066..3638,4068..5540,5652..6138))
	/note="gene_id:T803.2"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/product="transposon protein-like"
	/complement(join(10176..10508,10551..10817,10928..11178,11221..12021,12122..12617,12704..13313,13412..14198))
	/note="gene_id:T803.3"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/product="helicase-like protein"
	/complement(join(15838..15949,16034..16176,16277..16393,16512..16827))
	/note="gene_id:T803.4
	unknown protein"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(join(17978..18087,18184..18326,18423..18537,18639..18900,18992..19054))
	/note="gb AF34238.1
	gene_id:T803.5
	strong similarity to unknown protein"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(join(19768..19837,19928..20012,20096..20204,20385..20434,20533..20587,20666..20803,20987..21109))
	/note="emb CAB87733.1
	gene_id:T803.6
	similar to unknown protein"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(join(21775..22104,22193..22405,22491..22614,22735..22934,23012..23141,23223..23310,23406..23533,23626..23722,23887..23952))
	/note="gene_id:T803.7"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/product="replication protein A1-like"
	/complement(25983..26370)
	/note="gb AF18641.1
	gene_id:T803.8
	similar to unknown protein"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(join(28153..28638,28734..28824,28916..29062))
	/note="contains similarity to retroelement pol
	polyprotein
	gene_id:T803.9"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(31714..33511)
	/note="gene_id:T803.10
	unknown protein"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	35905..37542
	/note="contains similarity to retroelement pol
	polyprotein
	gene_id:T803.11"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/complement(join(39525..39821,39951..40199,40294..40432,40572..40790,41263..41644,41747..42020,42161..42415))
	/note="gene_id:T803.12"
	/codon_start=1
	/evidence=not_experimental
	/product="5-epi-aristolochene synthase-like protein;
	terpene synthase"
	/pseudo
	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB01981.1"
	/db_xref="GI:9294130"
	/translation="MAVARTVFGGLSTYLHQAPLFLKTSQSLEPPPSLSLKPWKDF VCWKATKTSSTSDLESGRPSILFSPISMGDYLSVSDSEDDIARETESYMKPYV RDLRSLSHNSNKKIRLIHLLISIGSYFESIEMLLNKAFELMIIAEEDDLFTI SIMFEVFRLYQHKMSCDSFVRKGEGRLEKSLVGVGRMLQLYQAAHGTSPDYIM







Gonzalez, A., Kremenetskaya, I., Kim, C., Lenz, C., Li, J., Liu, S.,  
 Lueros, S., Schwartz, J., Shin, P., Toriumi, M., Vyotskaya, V.,  
 Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R.W.  
 Direct Submission  
 Submitted (26-AUG-1999) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 3 (bases 1 to 108061)  
 Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,  
 Gonzalez, A., Kremenetskaya, I., Kim, C., Lenz, C., Li, J., Liu, S.,  
 Lueros, S., Schwartz, J., Shin, P., Toriumi, M., Vyotskaya, V.,  
 Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R.W.  
 Direct Submission  
 Submitted (11-DEC-1999) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 4 (bases 1 to 108061)  
 Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologos, A.  
 and Davis, R.W.  
 Submitted (07-APR-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 COMMENT  
 On Apr 7, 2000 this sequence version replaced gi:6560894.  
 Bases 1-4149 of IGF clone F2J6 overlap with bases 91042-95190 of  
 TMU clone T10P12 (gb|AC007203) and an unknown number of bases  
 overlap at the end of F2J6 overlap with IGF clone F28H19  
 (gb|AC006423).  
 FEATURES  
 source  
 1. 108061  
 /location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="F2J6"  
 complement(4642..5202)  
 /gene="F2J6.1"  
 complement(join(4642..4831,4916..4978,5051..5202))  
 /gene="F2J6.1"  
 /note="Unknown protein; Location of EST F2E8T7, gb|N96384"  
 /codon\_start=1  
 /protein\_id="AAF63116.1"  
 /db\_xref="GI:7523677"  
 /translation="MAMFVQIYVAGMPLSKVAISQGNPFVVFVROAFAALASP  
 FAPFLESKSSPLSLIKIFPISLGLGLSLNLYVALENTATTAATTAATNAPSI  
 IFVALFRITQSLSTLSTKPCDDPVFEVL"  
 26596..28759  
 /gene="F2J6.2"  
 join(26596..26754,26977..27024,27093..27167,27254..27315,  
 27401..27467,27581..27694,27790..27876,27954..28024,  
 28113..28178,28273..28359,28484..28559,28646..28759)  
 /gene="F2J6.2"  
 /codon\_start=1  
 /product="putative fructose 1,6-bisphosphatase"  
 /protein\_id="AAF63117.1"  
 /db\_xref="GI:7523678"  
 /translation="MDRAADAHRTDLMTTRFVINEOSKTPESGDTILLSHVLCG  
 KFYCSAVNKAAGLKLIGLAGEITNIGEEQRKLDVLSNDVFNALVSSGRFSVLSEED  
 EEAIFEPKSRGRKCVDFDLSDSSNIDCVSIGTIGIYTLDTDEPTADYLPEN  
 ENVAAGYCMGSSGLVLTGTVGHGFTLPSGLFEITLHPDIKIPKNGIYVNEGN  
 AOMNDGPTTYVEKCKRPKDGSPAKSLRYGSMVADVHRTLDYLGIFLYPDKKSPNG  
 KLRVLYEPFMSPLMEQAGGAFTGKRRLADIVPEKIHNSPIFGSYDVEEIKALY  
 AAEKKN"  
 30617..31261  
 /gene="F2J6.3"  
 join(30617..30834,30985..31261)  
 /gene="F2J6.3"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF63123.1"  
 /db\_xref="GI:7523684"  
 /translation="MSDELWNIDIOHMLGRELFIPIYHAYAGAFASNRLSHLGRILNPG

TOSVRAILELPYOMGLGSOVGHILDRGISHPSYSERTIEFTASTLGEVANDFNE  
 ETTQIFIRKVKVNDIEPTLRFRRVRVRESREGAMGNFNEIKIQOCTNCCRVNHQV  
 SHCP"  
 36931..37350  
 /gene="F2J6.4"  
 36931..37350  
 /gene="F2J6.4"  
 /note="Unknown protein"  
 /codon\_start=1  
 /protein\_id="AAF63118.1"  
 /db\_xref="GI:7523679"  
 /translation="MADHQEDDLKALKMSMOYNPEPKRSKPIEEEGSGSGG  
 ESPKARKRLQRELMAAAEKRWLPFKSPSPVAKAVLPISVGGDKDEEVGSCIG  
 KELSEESDOLFSTVFENEVSKSVLAQMTNGIRYDS"  
 39054..41001  
 /gene="F2J6.5"  
 join(39054..39133,39234..39345,39452..39628,40186..41001)  
 /gene="F2J6.5"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF63119.1"  
 /db\_xref="GI:7523680"  
 /translation="MGVWTLSDRHHVSDRSSTVEAKTKALVRSCEILPMCGDNN  
 RAVIASFTNEEDSSNQDEAMAGLPIESASDLQKILREFTTQASAKNLEGITL  
 AFQSRMGALLFLISALLSRGLEIYVIGSSEHYVLEADPDSVEENLELRESQIR  
 GFITLLESLNCKRGVONLKCPEKPIYVIGSSEHYVLEADPDSVEENLELRESQIR  
 RAPPAROOSGGGFGITVEAFHVOVQENIRIPKIKLDICATGFIWSELMOVLELD  
 ONLGICIDATGIMKKVPIYHFNIGIAKSDINGGCOMAYGAGVPMQRPRTLTANS  
 VPKWTEPEITCALPPSSSEKSEVQMPKVOHAPLVDICRTIWMSTRACWSDDPS  
 IV"  
 42764..44429  
 /gene="F2J6.6"  
 join(42764..43362,43458..43590,44048..44132,44221..44429)  
 /gene="F2J6.6"  
 /note="Contains bZIP transcription factors basic domain  
 protein motif."  
 /codon\_start=1  
 /product="putative transcription factor"  
 /protein\_id="AAF63120.1"  
 /db\_xref="GI:7523681"  
 /translation="MEGGGRGPNOITLSEIHMEAPRORISHRRARSETFSGESTI  
 DDLLPDSIDFSSLDLNPAPPOSOQOPQSPMSVSEETSSNGVVPNSLPK  
 PEARFGRRHVSFSDFFDLGVTKEEFTIATSGEKKGHHSRSMSDGESSAS  
 FNIESIIVASVSGSKKMGMDRLAEILLDPKRAKRIIAROSARSKERIRY  
 TGELEKRVOTIIONATITLSAQVTLQGTSELTENNHIKRLQALDQALRALNE  
 ALRDELNRKAYVAGEIIPQNGNSNTRAFSSQASAMNFGKTKQOMSTNGQPSLPST  
 MFTKRG"  
 complement(44946..46710)  
 /gene="F2J6.7"  
 complement(join(44946..45509,45595..45731,45811..46004,  
 46088..46319,46389..46710))  
 /gene="F2J6.7"  
 /codon\_start=1  
 /product="putative histidine decarboxylase"  
 /protein\_id="AAF63121.1"  
 /db\_xref="GI:7523682"  
 /translation="MVGSLSDQFLSMATLIEKLDIISDDPPTAVVTEPPVPTNG  
 IGAKGGGGGGRVYLGRIHTTSLAVTEPVNDEPFGDKRAYSVIARVTRKLYER  
 TKNLGYPRYNLDIFYGALGQLOHSINNLDGPTLESNTGVHSRVEGLVDFARLWE  
 IERSDGMYTINCQTEGNLHGLILVGREPGLIYASRESHYVFKAKARVMRCSEY  
 DLDNDGVCIDCDLRRKLLANLKVNDIIVKGVADVDLIVTILLECSCSHR  
 FYIHGQDLFEIAMPYKRAKVFENKIGSVSGHFGVCGCPMGOJLTREMHIV  
 LSNVEYLASRDATIMGSRNGHAPLFLMYTLNRRGCGYGFQKQECGLRMAYILKDRIR  
 EAGISAMINELSTVYFERPKDEEFYRMQJACQGDIAHVVMPSVTIEKLDNLFKDL  
 VKHRLINVEDSQPPCLASVEGTNCCIPARK"  
 complement(47997..49751)  
 /gene="F2J6.8"  
 complement(join(47997..48982,49058..49751))  
 /gene="F2J6.8"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF63122.1"  
 /db\_xref="GI:7523683"





and XhoI was ligated to modified lambda FIC-1 vector (Carlini et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

## FEATURES

Source

Location/Qualifiers  
1. 621  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="RAF109-06-F07"  
/clone\_lib="RAF19"  
/dev\_stage="plants at various developmental stages from germination to mature seeds"  
/lab\_host="DH10B"  
/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"  
BASE COUNT 154 a 146 c 147 g 173 t 1 others  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 612; DB 10; Length 621;  
Matches 612; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

73 GAAGTGAAGATGAGTGTGTAAGACAAACCGTATACCGTTGGTTCAGAGTCCG 132  
Db 1 GAAGTGAAGATGAGTGTGTAAGACAAACCGTATACCGTTGGTTCAGAGTCCG 60  
QY 133 AACGACACGACCGCTAAACGAAATCTCAAAATTCGCGATTCGTTGCTACTGT 192  
Db 61 AACGACACGACCGCTAAACGAAATCTCAAAATTCGCGATTCGTTGCTACTGT 120  
QY 193 TCAAAACCTGGTGGCGGAACCGAACTGTGCTTTTCGGTTCGGGTTTCGCTTT 252  
Db 121 TCAAAACCTGGTGGCGGAACCGAACTGTGCTTTTCGGTTCGGGTTTCGCTTT 180  
QY 253 CTTCGCACTCCGAGCGCTAGTTGGTTTATTTTCAGAGTTTCTCTCTTTTAT 312  
Db 181 CTTCGCACTCCGAGCGCTAGTTGGTTTATTTTCAGAGTTTCTCTCTTTTAT 240  
QY 313 CGGCGACGACGCTGATGTTCTGTCAAAAGCTTAACGATCCGATCGAGTGCAG 372  
Db 241 CGGCGACGACGCTGATGTTCTGTCAAAAGCTTAACGATCCGATCGAGTGCAG 300  
QY 373 TAAAGAGAAGACAGCGATTTGTGTAGATCGACGGCGAAGCTGTGATCCGTC 432  
Db 301 TAAAGAGAAGACAGCGATTTGTGTAGATCGACGGCGAAGCTGTGATCCGTC 360  
QY 433 ATCGAGGAGATACGTTTCATCGGTTTCGATCCAAATCGGAGATTTGAGATCT 492  
Db 361 ATCGAGGAGATACGTTTCATCGGTTTCGATCCAAATCGGAGATTTGAGATCT 420  
QY 493 ATCGAAATTCATTAATACATCTCCAAATCTTCTGAAGAGTCCAAATCCGATCT 552  
Db 421 ATCGAAATTCATTAATACATCTCCAAATCTTCTGAAGAGTCCAAATCCGATCT 480  
QY 553 CACCACTACTGTAACCGCGGCTGATTAAGTCCGCGATTTCAAAATTAATCCGATCT 612  
Db 481 CACCACTACTGTAACCGCGGCTGATTAAGTCCGCGATTTCAAAATTAATCCGATCT 540  
QY 613 CGGCGATTCGAATGCGAGCTGAGATCTGGGTTTTCGATTCAGGATTAATCTTGT 672  
Db 541 CGGCGATTCGAATGCGAGCTGAGATCTGGGTTTTCGATTCAGGATTAATCTTGT 600  
QY 673 CACTCTTTGAAA 684  
Db 601 CACTCTTTGAAA 612

RESULT 2  
AV548305/c AV548305 556 bp mRNA linear EST 06-SEP-2000  
DEFINITION AV548305 Arabidopsis thaliana roots Columbia Arabidopsis thaliana

CDNA clone R2L51d11F 3', mRNA sequence.  
AV548305  
VERSION AV548305.1 GI:8719718  
KEYWORDS EST  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

REFERENCE  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries  
JOURNAL DNA Res. 7, 175-180 (2000)  
MEDLINE 20363093  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: [asamizu@kazusa.or.jp](mailto:asamizu@kazusa.or.jp), URL: <http://www.kazusa.or.jp/en/plant/>.

## FEATURES

Source

Location/Qualifiers  
1. 556  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="R2L51d11F"  
/clone\_lib="Arabidopsis thaliana roots Columbia"  
/tissue\_type="roots"  
/note="Vector: plasmid pUC19 SK-; Site 1: EcoRI; Site 2: XhoI"  
BASE COUNT 171 a 129 c 103 g 153 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 556; DB 10; Length 556;  
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1823 ATGCTACTGAACCAATCAACGAGCCATTATTAAGAACTGTAGAGCAAGAGCTA 1882  
Db 556 ATGCTACTGAACCAATCAACGAGCCATTATTAAGAACTGTAGAGCAAGAGCTA 497  
QY 1883 TGTAGAGATCAGAAATTTCTTATGCTCAGATGGGAGCGGCTTCCGATAGAG 1942  
Db 496 TGTAGAGATCAGAAATTTCTTATGCTCAGATGGGAGCGGCTTCCGATAGAG 437  
QY 1943 CTGAATCTCAAACTCAACTTTTGTGATCTACAAATAGTGTGAAGAGATCTACTGCTG 2002  
Db 436 CTGAATCTCAAACTCAACTTTTGTGATCTACAAATAGTGTGAAGAGATCTACTGCTG 377  
QY 2003 GTGTTCTGAGAGCTGTGGATTTGATGCCATATTTGCAATCACTTTAGGGATTC 2062  
Db 376 GTGTTCTGAGAGCTGTGGATTTGATGCCATATTTGCAATCACTTTAGGGATTC 317  
QY 2063 CCAACCTGACCGAGCAGTGGATTCGACAAATGTTTGGCTTGTGGTGAAGAGATC 2122  
Db 316 CCAACCTGACCGAGCAGTGGATTCGACAAATGTTTGGCTTGTGGTGAAGAGATC 257  
QY 2123 CACATGCGGTTTGGCTTAGAAGTGTGTATCCACGAAACCAATGATTAATCTTCAGGCGTTT 2182  
Db 256 CACATGCGGTTTGGCTTAGAAGTGTGTATCCACGAAACCAATGATTAATCTTCAGGCGTTT 197  
QY 2183 CATCAATTCACCTTGAGTAACCAACATTTGTTCACTGCTCAATTAATTAAGTGCCTACCA 2242  
Db 196 CATCAATTCACCTTGAGTAACCAACATTTGTTCACTGCTCAATTAATTAAGTGCCTACCA 137  
QY 2243 AGTTGGTTGAGTATACGTTTGTGATATAGACTTGGTGTAAATTTCTGGTGAAGC 2302  
Db 136 AGTTGGTTGAGTATACGTTTGTGATATAGACTTGGTGTAAATTTCTGGTGAAGC 77  
QY 2303 ATTTTATACCAATTTGTAAGGCTTTTAACTCTTGGAATACTGGGGAATAATAATAA 2362  
Db 76 ATTTTATACCAATTTGTAAGGCTTTTAACTCTTGGAATACTGGGGAATAATAATAA 17

QY 2363 GTTGATTTCAATCTT 2378  
 |||  
 DB 16 GTTGATTTCAATCTT 1

RESULT 3  
 AV556975/c 486 bp mRNA linear EST 06-SEP-2000  
 LOCUS AV556975 Arabidopsis thaliana green siliques Columbia Arabidopsis  
 DEFINITION thaliana cDNA clone S0057109F 3', mRNA sequence.  
 ACCESSION AV556975  
 VERSION AV556975.1 GI:8728390  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 486)  
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 DNA Res. 7, 175-180 (2000)  
 20363093

JOURNAL  
 MEDLINE  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yama 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
 Location/Qualifiers

FEATURES  
 source 1..486  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="S0057109F"  
 /clone\_lib="Arabidopsis thaliana green siliques Columbia"  
 /tissue\_type="green siliques"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 155 a 114 c 91 g 126 t  
 ORIGIN

Query Match 20.3%; Score 486; DB 10; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-108; Indels 0; Gaps 0;  
 Matches 486; Conservative 0; Mismatches 0;

QY 1893 CAGAAATTTATGCGTCAGATGGTGAGCGGCTAGCGATAGAGCCTGAATCTCA 1952  
 |||  
 DB 486 CAGAAATTTATGCGTCAGATGGTGAGCGGCTAGCGATAGAGCCTGAATCTCA 427

QY 1953 AACTCACTTTTGATTTACATGAGTCTGTAGAGAGTTTACTTGGTGTCTCTGG 2012  
 |||  
 DB 426 AACTCACTTTTGATTTACATGAGTCTGTAGAGAGTTTACTTGGTGTCTCTGG 367

QY 2013 AACTGTGATTTGATTCATATTTGCAATCACTTGAAGGGATTCGGCACCACAACTGAC 2072  
 |||  
 DB 366 AACTGTGATTTGATTCATATTTGCAATCACTTGAAGGGATTCGGCACCACAACTGAC 307

QY 2073 CCAGCATGAGATTCGCACAAATGTTTGGCTTGTGTGAGAGAGATCCACATGGCGT 2132  
 |||  
 DB 306 CCAGCATGAGATTCGCACAAATGTTTGGCTTGTGTGAGAGAGATCCACATGGCGT 247

QY 2133 TTGCTAGAGAAATGGTGTATCCAGAACCATGATTTACTTCAAGCGTTTCATCAATCA 2192  
 |||  
 DB 246 TTGCTAGAGAAATGGTGTATCCAGAACCATGATTTACTTCAAGCGTTTCATCAATCA 187

QY 2193 CCTTGAGTAACAACATTTTTCAGTGTCCCAATTTAGTGGGTGACCAAGTTGGTGG 2252  
 |||  
 DB 186 CCTTGAGTAACAACATTTTTCAGTGTCCCAATTTAGTGGGTGACCAAGTTGGTGG 127

QY 2253 AGTAACTGTTTGCATATAGACTGGGTGTAAATTTCTTGGTGTAAACATTTTATAC 2312

DB 126 AGTAACTGTTTGCATATAGACTGGGTGTAAATTTCTTGGTGTAAACATTTTATAC 67  
 |||

QY 2313 CCATTTAAGGCTTTAACTCTTGGAAGAACTTCGGGAAATTAATTAAGTGATTTCA 2372  
 |||  
 DB 66 CCATTTAAGGCTTTAACTCTTGGAAGAACTTCGGGAAATTAATTAAGTGATTTCA 7

QY 2373 AATCTT 2378  
 |||  
 DB 6 AATCTT 1

RESULT 4  
 A1995676/c 475 bp mRNA linear EST 08-SEP-1999  
 LOCUS A1995676  
 DEFINITION 701516116 A. thaliana, Columbia Col-0, inflorescence-2 Arabidopsis;  
 thaliana cDNA clone 701516116, mRNA sequence.  
 ACCESSION A1995676  
 VERSION A1995676.1 GI:5842581  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 475)  
 Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroun, B., Gilliland, D.,  
 Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Broszka, P.,  
 Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.,  
 Rose, M., Warren, B., Ton, B., Kasuriy, K., Borillo, C., Carpilo, T.,  
 Pollick, J., Suzuki, G., Argentine, C., Shah, S., Nobrigha, A., Murry, L.,  
 Turner, C., Krikorian, S., Elder, L. and Hanson, D.  
 Arabidopsis thaliana Gene Expression Microarray  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: David Smoller, Ph.D.  
 Genome Systems, Inc., a wholly owned subsidiary of Incyte  
 Pharmaceuticals, Inc.  
 4633 World Parkway Circle, St. Louis, MO 63134, USA  
 Tel: 877-577-2733  
 Fax: 314-427-3324  
 Email: service@genomesystems.com.

FEATURES  
 source 1..475  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia Col-0"  
 /db\_xref="taxon:3702"  
 /clone="701516116"  
 /clone\_lib="A. thaliana, Columbia Col-0, inflorescence-2"  
 /tissue\_type="inflorescence"  
 /dev\_stage="4 - 7 weeks"  
 /note="Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; cDNA  
 library was derived from untreated inflorescence tissue.  
 from Arabidopsis thaliana, Columbia Col-0, at 4 - 7  
 weeks. Plants were grown in 1:1:1 peat  
 moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C  
 under constant light, and watered with fertilizer. cDNA  
 synthesis was initiated using a NotI-oligo(dT) primer.  
 Double-stranded cDNA was blunt-ended, ligated to SalI adaptors  
 , digested with NotI, size-selected, and cloned into the  
 NotI and SalI sites of the pSPORT vector."

BASE COUNT 149 a 118 c 90 g 118 t  
 ORIGIN

Query Match 19.8%; Score 473.4; DB 9; Length 475;  
 Best Local Similarity 99.8%; Pred. No. 2e-105; Indels 1; Gaps 0;  
 Matches 474; Conservative 0; Mismatches 1;

QY 1874 GAGAACTATGTTGAGATCAGAAATTTCTATGCGTGAAGTGGGCGGCTAGCGTTC 1933  
 |||  
 DB 475 GAGAACTATGTTGAGATCAGAAATTTCTATGCGTGAAGTGGGCGGCTAGCGTTC 416

QY 1934 CGATAGAGCCTGAATCTCAACTCAACTTTTGATTTCAATGAGTCTGAAGAGTTTC 1993  
 |||

	Query Match	Similarity	Score	DB	Length
Best Local	437	100.0%	18.2%	10	437
Matches	437	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

  

	Query	Match	Similarity	Score	DB	Length
1	GCCAAATGACAGGCGTTGTGTAGTACAAATCAGCGTTTACGCGATTTGTATACCAAT	809	100.0%	18.2%	10	437
	1	GCCAAATGACAGGCGTTGTGTAGTACAAATCAGCGTTTACGCGATTTGTATACCAAT	60			

Query Match	17.8%	Score 425.4	DB 9	Length 728
Best Local Similarity	76.1%	Pred. No. 1.1e-93		
Matches 554	Conservative	0	Mismatches 162	Indels 12
			Gaps	2
OY	784	CGGTTTACGCATTGTGAAGCCATCAACGAAGAAGTCACGCTGAAAGTTGGCGATGG	843	







OY	1799	TTACTTCTGAAAGGGGTGTACATGCTACTGAAACCAATCAACGAACCACTTATTAAG	1858
Db	361	TTAAACCGACAAAGTGGATGGAGGAGTGACTGAGCTTATCAAAACCAAGTTGTCAAAG	420
OY	1859	AACTCTTAGAGGCAAGAAACCTATGTGGAGATACGAATTCCTATGCGTCAGATGGGTG	1918
Db	421	TATTCTTAAGGCAAGGAGGTATTATGCTTGAGATCAGAAATACATCGGAACATGGGAG	480
OY	1919	AGCGGGCTAGCGTTCCTCATAGCCTGATATCCAATCAACTTTGGATTCTACATGA	1978
Db	481	AGGAGAGCAATGTGTTCCGATAGAACCCGATACACAGACTAAACTTTTGGATCTCTACTGA	540
OY	1979	GTGCGAAGGAGATCTACTTCTGGTGTTCCGAGCGGGAGATTATGCCAATATTG	2038
Db	541	ATATGGAAGGAGTTTGTGTGGCCCGAGTTCTCGACAGCATGAGATTATGACAGTCTTTT	600
OY	2039	CAATCACTTTAGGGGATTCGGGACACAAATGACCACAGCATGAGTTCGGACATGTTT	2098
Db	601	CTGTACCTTGGGTGATCTCTAGCAAGAAGCGTAGCAAAATGCAATGGAGTTCGGACATGTTT	660
OY	2099	TGGCTTTTGGTGAGAAAGATCCACATGGCGTTTGCTTGAAAAGTGTGATCCAGCA	2158
Db	661	TGGCTTTTGTAGTAGGGAAGAACCTCCAAAGTGTGTTGCTTAAAAAGTGGCATCCACGAT	720
OY	2159	CCACATGTATTACTCAGGCGTTTCATCAATTCACCTTGAGTAAG	2204
Db	721	GCCTGGAAATCATGTGTCCGAGTTCCTTAGATTAACATCAAGTAGAC	766

RESULT	10
AV788756/c	
LOCUS	417 bp mRNA linear EST 28-MAR-2002
DEFINITION	AV788756 RABF16 Arabidopsis thaliana cDNA clone RAFL06-80-O03 3'
	mRNA sequence.
ACCESSION	AV788756
VERSION	AV788756.1 GI:19807546
KEYWORDS	EST.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana

REFERENCE	1 (Pages 1 to 417)
AUTHORS	Seki, M., Natusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Morimatsu, M., Hayashizaki, Y. and Shinozaki, K.
TITLE	Large scale analysis of Arabidopsis full-length cDNA (2002p)
JOURNAL	Unpublished (2002)
COMMENT	Contact: Motoaki Seki

FEATURES  
SOURCE

Email: [mesekertc.riken.go.jp](mailto:mesekertc.riken.go.jp)  
An Arabidopsis full-length cDNA library was constructed, essentially as reported previously (Seki et al., 1998). This clone is in a modified pBluescript vector as a SstI/XhoI insert. Please visit our web site ([http://www.gsc.riken.go.jp/e/Plant/index\\_e.html](http://www.gsc.riken.go.jp/e/Plant/index_e.html)) for further details.

Location/Qualifiers  
1. .417

```

/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFLO6-80-003"
/clone_lib="RAFLO6"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DHIO6"
/note="Site_1: Sstr; Site_2: XhoI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24

```

	hr) treatments"				
BASE COUNT	136 a	96 c	76 g	109 t	
ORIGIN					
Query Match		16.9%	Score 406;	DB 10;	Length 417;
Best Local Similarity		99.8%;	Pred. No. 7e-89;		
Matches 417;	Conservative	0;	Mismatches	0;	Indels 1; Gaps 1;

OY	1959	ACATTTGGATCTTCACATGAGTCTGTAAGAGATTCACATGGCGGTTCTCGAGCTGG	2018
Db	417	ACTTTTGGATCTTCACATGAGTCTGTAAGAGATTCACATGGCGGTTCTCGAGCTGG	358
OY	2019	TGATTTGGATGCCATATTTCGAATCACTTTAGGGGATTCGGGACCAAACTGACCCAGGC	2078
Db	357	TGATTTGGATGCCATATTTCGAATCACTTTAGGGGATTCGGGACCAAACTGACCCAGGC	298
OY	2079	ATGAGATTCGCACACATGTTTTGGCCTTGTTGGTGAGAGAAATCCACATGGCGTTGGCT	2138
Db	297	ATGAGATTCGCACACATGTTTTGGCCTTGTTGGTGAGAGAAATCCACATGGCGGCT	239
OY	2139	AGAAAGTGGTGATCCACGAAACCACTGATATTACTTGAGGCGTTTCATCAATCACTCTGA	2198
Db	238	AGAAAGTGGTGATCCACGAAACCACTGATATTACTTGAGGCGTTTCATCAATCACTCTGA	179
OY	2199	GTAACACACATGTTGTTCACTGCTCCAAATTATTAGTGCCTGCACCAAGTTGCGTTGAGTATA	2258
Db	178	GTAACACACATGTTGTTCACTGCTCCAAATTATTAGTGCCTGCACCAAGTTGCGTTGAGTATA	119
OY	2259	CTGTTTGGCATATGACTTGGGTCGCTAAATTTCTTGCTGAAGCATTTTATACCATTG	2318
Db	118	CTGTTTGGCATATGACTTGGGTCGCTAAATTTCTTGCTGAAGCATTTTATACCATTG	59
OY	2319	TAAAGCTTTAACCTTGGAACCTGGCGGGAATAATAAATTAAGTGTGATTTCAATC	2376
Db	58	TAAAGCTTTTAACCTTGGAACCTGGCGGGAATAATAAATTAAGTGTGATTTCAATC	1

AV796045/c	430 bp	mRNA	linear	EST 29-MAR-2002
LOCUS	AV796045	AV796045	Arabidopsis	thaliana cDNA clone RAFL09-06-F07 3'
DEFINITION	AV796045	RAFL09	Arabidopsis	thaliana cDNA clone RAFL09-06-F07 3'
ACCESSION	AV796045	RAFL09	Arabidopsis	thaliana cDNA clone RAFL09-06-F07 3'
VERSION	AV796045.1	GI:19830028	EST	thale cress.
KEYWORDS				
SOURCE				

REFERENCE  
AUTHORS  
Oono, Y., Tanusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,  
Sakurai, T., Carrington, P., Kawai, J., Itoh, M., Ishii, Y.,  
Arakawa, F., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.  
and Shimizuaki, K.  
Large scale analysis of Arabidopsis full-length cDNA (2002b)  
Unpublished (2002)  
Contact: Motoaki Seki

```

FEATURES
SOURCE
    Email: mskei@rtc.riken.go.jp
    An Arabidopsis full-length cDNA library was constructed essentially
    as reported previously (Sei et al., 1998). cDNA cleaved with BamHI
    and XhoI was ligated to modified Lambda Ptc-1 vector (Garniuel et
    al., submitted for publication) digested with BamHI and SalI. This
    clone is in a modified pluescript vector. Please visit our web
    site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
    details.
    Location/Qualifiers
        1..430

```

/organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /clone="RAF109-06-F07"  
 /clone\_1lb="RAF19"  
 /dev\_stage="plants at various developmental stages from germination to mature seeds"  
 /lab\_host="DH10B"  
 /note="Site\_1: BamHI; Site\_2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

BASE COUNT 139 a 98 c 80 g 113 t

ORIGIN

Query Match 16.7%; Score 401.2; DB 10; Length 430;  
 Best Local Similarity 98.8%; Pred. No. 1.1e-87;  
 Matches 425; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1949 CTCGAACCTGATTTGATCTCAATAGATGCTGAAGAGATCTACTT-GCTGGTGT 2007  
 |||||||  
 Db 430 CTCGAACCTGATTTGATCTCAATAGATGCTGAAGAGATCTACTTGGCTGGTGT 371

QY 2008 CCTGAGCTGATGATTTGATCCATATTTGCAATCACTTTAGGGGATTCGGGACCAAA 2067  
 |||||||  
 Db 370 CCTGAGCTGATGATTTGATCCATATTTGCAATCACTTTAGGGGATTCGGGACCAAA 311

QY 2068 CTGACCCAGCAGTGCACAAATGTTGGCTTTGGTGTGAGAGAAGATCCACAT 2127  
 |||||||  
 Db 310 CTGACCCAGCAGTGCACAAATGTTGGCTTTGGTGTGAGAGAAGATCCACAT 251

QY 2128 GGGCTTTGCCAGAAATGCGATCCAGAACACATATTAATCTTACCTGAGCGTTTCATCA 2187  
 |||||||  
 Db 250 GGGCTTTGCCAGAAATGCGATCCAGAACACATATTAATCTTACCTGAGCGTTTCATCA 191

QY 2188 ATTACCTTGAGTAACAACATGTTTCAGTGTCCAAATATTA-GTGGCGACCAAGTT 2246  
 |||||||  
 Db 190 ATTACCTTGAGTAACAACATGTTTCAGTGTCCAAATATTAAGTGTCCGACCAAGTT 131

QY 2247 CGGTGAGTACTGTTTGCATATAGACTGGTGTCTAAATTTCTGGTGAAGCATTT 2306  
 |||||||  
 Db 130 CGGTGAGTACTGTTTGCATATAGACTGGTGTCTAAATTTCTGGTGAAGCATTT 71

QY 2307 TTATACCATTTGATAGCTTTAATCTGTGAAAACTTGGGGAATAAATAAAGTTG 2366  
 |||||||  
 Db 70 TTATACCATTTGATAGCTTTAATCTGTGAAAACTTGGGGAATAAATAAAGTTG 11

QY 2367 ATTCAATC 2376  
 |||||||  
 Db 10 ATTCAATC 1

RESULT 12  
 BQ118666/c 844 bp mRNA linear EST 17-Apr-2002  
 LOCUS BQ118666  
 DEFINITION EST604242 mixed potato tissues Solanum tuberosum cDNA clone s1MEC81  
 3' end, mRNA sequence.  
 ACCESSION BQ118666  
 VERSION BQ118666.1 GI:20170628  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 1 (bases 1 to 844)  
 Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
 Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and  
 Karimicheva, S.A.  
 Generation of a set of potato cDNA clones for microarray analyses  
 unpublished (2002)  
 CONTACT Robin Buell  
 The Institute for Genomic Research  
 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Email: potatoc@tigr.org

This clone is available through the Research Genetics, contact the  
 Research Genetics for further information 1-800-711-6195 or  
 cna@resgen.com  
 Seq primer: F7.  
 Location/Qualifiers  
 1..844  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec or Blinje"  
 /db\_xref="taxon:4113"  
 /clone="s1MEC81"  
 /clone\_1lb="mixed potato tissues"  
 /tissue\_type="mixed tissues"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; supplier: Combination of untreated and Phytophthora  
 infestans-treated libraries of stolons, leaves, leaflets,  
 axillary buds of stem explants, petioles, germinating eyes  
 , tubers, or roots."

BASE COUNT 229 a 215 c 155 g 245 t

ORIGIN

Query Match 15.7%; Score 377.2; DB 14; Length 844;  
 Best Local Similarity 67.5%; Pred. No. 7.6e-82;  
 Matches 560; Conservative 0; Mismatches 268; Indels 2; Gaps 2;

QY 1316 AAGGAAGTTTGGCTTCTGATCTAGATGTTATCATATGATAGCAACAGTCTCAT 1375  
 || |||||  
 Db 838 AAGGAAGTTTGGCTTCTGATCTAGATGTTATCATATGATAGCAACAGTCTCAT 779

QY 1376 GTCTTGCAAGAGAGAGTGGGAAGTGGTGTGATGTCAGCTGTGCTCATAGGAAGTC 1435  
 || |||||  
 Db 778 GCATTTGCAAGAGAGTGGGAAGTGGTGTGATGTCAGCTGTGCTCATAGGAAGTC 719

QY 1436 AGCGTTATGTTGGCTTCTGATCTAGATGTTATCATATGATAGCAACAGTCTCAT 1495  
 || |||||  
 Db 718 AACCTATATTCGGTGTGCTGATAGTGTCTTCTGCGCAGAAATGCAAGTATGGCA 659

QY 1496 TGCCATTAA-ATGAGTTATTTGTAACAATTTGAAGGGAATGGAACAATAGAGAACT 1554  
 || |||||  
 Db 658 CACCACATAATGGAAGTATGATGATGCTTAAGAACAAAGTGGAGCAATGAGAGACC 599

QY 1555 GAGTTTCTTTACACACAGATGATATCTTTCTGAGAGAACTGGAAGTGGGATTC 1614  
 || |||||  
 Db 598 AAGTTTCTTTACACACAGATGATATCTTTCTGAGAGAACTGGAAGTGGGATTC 539

QY 1615 TCCACA-CCATCAATGATAGTGCAGTAAGAAAGTGGCAATGCTGATCCAGAGAGGC 1673  
 || |||||  
 Db 538 TCAACCCCATCATATGTTGAGCTGTTAAGAAATGCAAGAGTCCGACCTCCAGATTC 479

QY 1674 ACGAANAATGCGAGAAATTTGCAATGCAATTTAGAACTGGAATTAAGCTTAAGCA 1733  
 || |||||  
 Db 478 TCTAGAAAATGGAAGTGTGTCAGAAAGAACTGCTCTTGAACCGCATCTAAATAC 419

QY 1734 TCTGAGAAATTAAGCTTAAGAAAGTGTGTCAGAAAGTGTGTCAGAAAGTGTGTCAG 1793  
 || |||||  
 Db 418 CTTAAGTAATTTGCAAGAGAGACATTTACAAAGCTTTAATGATGATCAATGCTCTCAG 359

QY 1794 TGTGCTTACTTCTGAAAAGTGGGTGTACATGCTACTGAAACCAATCAAGAGCATAT 1853  
 || |||||  
 Db 358 CCGTCTTCTGCTGTAAGAGTGTGTCAGAAAGTGTGTCAGAAAGTGTGTCAGAAAGTGT 299

QY 1854 TAAAGAACTTTAGAGCAAGAGAGAGTATGTTAGAGTCAAGATTTTATAGCTCAGAT 1913  
 || |||||  
 Db 298 TAAAGAACTTTAGAGCAAGAGAGTATGTTAGAGTCAAGATTTTATAGCTCAGAT 239

QY 1914 GGGTGAAGGCGCTAGCTGATGAGAGCCGATGCAAACTGGAATCTTTGCTTCTATC 1973  
 || |||||  
 Db 238 GGGTGAAGGCGCTAGCTGATGAGAGCCGATGCAAACTGGAATCTTTGCTTCTATC 179

QY 1974 AATGAGTCTGAGAGAGTCTACTGCTGTGTTCTGAGAGTGTGGAATTTGATGTCAT 2033  
 || |||||  
 Db 178 AATGAGTCTGAGAGAGTCTACTGCTGTGTTCTGAGAGTGTGGAATTTGATGTCAT 119

OY	2034	ATTTCACATGACTTTAGGGATTTCCCGGCACCACAACTGCCAGCAGCATGTTCACA	2093
DB	118	TTTTGCAGTACCCTCGGGGCTTCAAGCCAAAATGTACTCAAGCTTGAGTTCCTTA	59
OY	2094	TGTTTTGGCTTGTGTTGTGAGAGAATCCATCCATGGCTTTGCCCTAATA	2143
Db	58	TGTTCTGCAATGCTAGTAGAGAGATTCATCGTGTGTCTCGAANA	9
RESULT 13			
Locus	AA042480/c	457 bp	mRNA linear EST 29-SEP-1997
DEFINITION	24833 CD4-16 Arabidopsis thaliana cDNA clone H9F577,		mRNA sequence.
ACCESSION	AA042480		
VERSION	AA042480.1 GI:2446214		
KEYWORDS	EST.		
SOURCE	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes I; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi s. I (bases 1 to 457) Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh, L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel, J.E. and Somerville,C. Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994) 95148729		
REFERENCE	AUTHORS		
JOURNAL MEDLINE	COMMENT		
	On Sep 29, 1997 this sequence version replaced gi:1520654. Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-P&L, Michigan State University,Plant Biology Bldg..E. Lansing,MI Tel.: 517-353-0854 Fax: 517-353-9168 Email: 22313cne@lm.cl.msu.edu Seq primer: T7.  Location/Qualifiers 1..457 /oranism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone_H9F577" /clone_11b-"CD4-16" /tissue_type="seedling hypocotyl" /dev_stage=""3 day-old" /note="vector: pbluescript SK-, Site.1: EcoRI; Site.2: EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old Arabidopsis thaliana (Colombia) seedling hypocotyls as template and oligo d(Tt) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/Noti adapters (Pharmacia) were ligated to each end. The cDNA was purified from unligated adapters by spun-column chromatography using sephacryl s-300 and size-fractionated on a 1% low melting point mini-gel. Size selected cDNAs (3 - 6 kb) were removed from the gel using agarose (New England Biolabs), phenol:chloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size fraction (0.1 ug) was co-precipitated with 1 ug of lambdaZapIT (Stratagene) digested, dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using GigaPack II gold packaging extract (Stratagene). We have determined that although first strand cDNA synthesis was initiated using dT, almost all of the cDNA begin 8-10 bp from the poly-A tail. The reason for the loss of the poly-A tail is most		

BASE COUNT	136 a	104 c	95 g	106 t	16 others
ORIGIN					
Query Match	14.7%	Score 353.2;	DB 9;	Length 457;	
Best Local Similarity	92.4%;	Pred. No. 6.4e-76;			
Matches 389;	Conservative 0;	Mismatches 30;	Indels 2;	Gaps 2;	
likely due to lower than anticipated nucleotide levels during the Klenow repair of ragged ends before the addition of linkers (3'-5' exo instead of 5'-3' pol). When this library is used please reference the ABRF and: Kieber, J. et al. (1993) Cell 72:427-441."					
OY 1927	AGCGCTCCGATTTGAGACCCGGAATCTCAAACTCAACTTTGGATTCTACATAGAG-TGCTGA	1985			
DB 420	AACGTCGCCAANAGNCNNTGATCNC-AACTCACTTTTGGATTCAACCATGCTTCTGA	362			
OY 1986	AGGACTTCTACTTGTCTGTGTTCTCTGAGCTGTGTGATTTGATTCATATTTGCATCAG	2045			
DB 361	AGGGATTTCTAACTTGTCTGTGTTCTCTGAGCTGTGTGATTTGATTCATATTTGCATCAG	302			
OY 2046	TTTACGGGATTCGGGACCAACTACCCAGGATGAGATTGCGCAATGTTTGGCTT	2105			
DB 301	TTTACGGGATTCGGGACCAACTACCCAGGATGAGATTGCGCAATGTTTGGCTT	242			
OY 2106	GTTGGTGAAGAGAAATTCACATGGGCTTGGCTTAAAGTGGTGAATTCACGAAACACATG	2165			
DB 241	GTTGGTGAAGAGAAATTCACATGGGCTTGGCTTAAAGTGGTGAATTCACGAAACACATG	182			
OY 2166	TATTTACTCAGCGCTTTTCATCAATTCACCTTGAGTAAACATCATGTTTCACTGTCAT	2225			
DB 181	TATTTACTCAGCGCTTTTCATCAATTCACCTTGAGTAAACATCATGTTTCACTGTCAT	122			
OY 2226	TATTTAGTGTCGCTACCAAGTTGGTTGAGTATACGTTTTCATATPAGACTTGGTCTA	2285			
DB 121	TATTTAGTGTCGCTACCAAGTTGGTTGAGTATACGTTTTCATATPAGACTTGGTCTA	62			
OY 2286	AATTTCTGGTGTAGACATTTTATACCATGTGAAGGCTTAACTCTTGGAAACCTTG	2345			
DB 61	AATTTCTGGTGTAGACATTTTATACCATGTGTAAAGCTTTAACTCTTGGAAACCTTG	2			
OY 2346	C 2346				
DB 1	C 1				
RESULT 14					
LOCUS	BG445135				
DEFINITION	GA_Ea0026P06f Gossypium arboreum 7-10 dpa fiber library Gossypium	841 bp	mRNA	linear	EST 15-MAR-2001
ACCESSION	BG445135				
VERSION	BG445135				
KEYWORDS	BG445135.1 GI:13354787				
SOURCE	EST.				
ORGANISM	Gossypium arboreum.				
REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
AUTHORS	Wing, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry				
TITLE	'D', Wood, T.C., Leslie, A. and Wilkins, T.A. An integrated analysis of the genetics, development, and evolution of the cotton fiber				
JOURNAL	unpublished (2000)				
COMMENT	Contact: Wing RA Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: TAATACGACTCACTATAGG High quality sequence stop:653.				

FEATURES	location/Qualifiers
source	1. 841 /organism="Gossypium arboreum" /strain="AKA" /cultivar="8400" /db_xref="taxon:29729" /clone="GA_Ea0026P06f" /clone_id="Gossypium arboreum 7-10 dpa fiber library" /tissue_type="Fibers isolated from bolls harvested 7-10 dpa" /lab_host="E. coli" /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	246 a 177 c 161 g 255 t 2 others
ORIGIN	
Query Match	14.7% Score 352.4; DB 12; Length 841;
Best Local Similarity	76.6% Pred. No. 9.3e-76;
Matches 469; Conservative	0; Mismatches 136; Indels 5; Gaps 3;
QY	681 GAAATGCGCTGTTGTTGCTTCCTGCTCGTGGAAAGTTTGGATGAGCTGAGCGCTACCTGCT 740
DB	185 GACATGGAAGTCGTTGCTTCGTCGCGGAAAGGTTTGGATGAGCTGCGGGCATATTAT 244
QY	741 ACTCGAAGACCAATGACAGGCGCTGTGTGAGTACAAATGACAGGTTTACCGCATGT 800
DB	245 TTTGGAGAGACCAATGCGTGGGATGTCTGAGTACAAAGTCGACGCTTTTATAGCATTTGT 304
QY	801 GAGCCAAATCAACGAGAAGTCAACGCTGAAAGTTGGCGCATGGAATGACAGATGTCA 860
DB	305 TAGGCCAATTCATGAAGCATATCAACCTGAAAGCTGGCGATGCTGTGGACCGATGTCA 364
QY	861 ATTAACATCACCACAGCTCTCGAGAAAGCATGTATAACTGTCACTGAATCATTTGAC 920
DB	365 GCTAACATCTCCTCAGCTTTCGAGGAAAGCATGTATAATGTCTCGGAAACATTTAAC 424
QY	921 TCTTCAGTCTGCTGCGAAGTGAATTCAGAAAGCCCTTGTAGAGCATGCGATACAGA 980
DB	425 ACTTCAGTGTGTATCTTCAAGTGAATCAAGAACCCCTTTGTAGAAATGCTGTTCAATA 484
QY	981 TGCATAGCTGCTGCTCATTTGGCA--ACGAGAAGGACAAAGATCATTTGCACAAACT 1037
DB	485 TACATAGCAGCTGCACATGCMACTTTGACAAAGATTAAGAAAGGCGCATTTAGATAACT 544
QY	1038 CTTATGGAAGGCTGTGATATACAAATNTAGGCTCCATGACTTTTACTCATTCGGA 1097
DB	545 ACTCTTACAGGCTCTTGATATTACACTTTAGGTTGCATGACTTTTACTCATTCAGGA 604
QY	1098 CCAGATAGAAATCGCTGGGCTTCATTTGACACACAGAAATGCTGGGCTACCCCTGAC 1157
DB	605 TCAGATAGAGACACTTGGCTCTCCGTTGACACCTGAGAGATTGGCTACTCTACACCGGT 664
QY	1158 TGCATCATCACATTCATGCTGCGGAGTCAAAATGCTGCTAATTCACAGCTGAAGTAC 1217
DB	665 TACATCATATTCAATTCATGCTGAGGAATCAAAATGAGACAAATTCACAAACCGAAGTTG- 723
QY	1218 AAAAATGCGCTAGGTTCTTCTGACAGCATGACACAG-CTGTGGTTGGAGCTGTAC 1276
DB	724 AAAAATGAATGGGGCTATCTGACAGCAATGACCACTGCCCGAANTGGCTGCTTAAC 783
QY	1277 ATTATCTTGAG 1288
DB	784 ATTACCTGGGG 795
RESULT 15	
B0118665	
LOCUS	B0118665 769 bp mRNA linear EST 22-JUL-2002
DEFINITION	EST604241 mixed potato tissues Solanum tuberosum cDNA STMEC81
ACCESSION	B0118665
VERSION	B0118665.2 GI:21919032
KEYWORDS	EST.
SOURCE	potato.

ORGANISM	Solanum tuberosum
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
AUTHORS	1 (bases 1 to 769) Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamchewa,S.A.
TITLE	Generation of a set of potato cDNA clones for microarray analyses
JOURNAL	Unpublished (2002)
COMMENT	On Apr 17, 2002 this sequence version replaced gi:20170627. Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potatodir.org This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com Seq primer: T3.
FEATURES	Location/Qualifiers
source	1..769 /organism="Solanum tuberosum" /cultivar="Kennebec or Binjle" /db_xref="taxon:4113" /clone_STMEC81 /clone_id="mixed potato tissues" /tissue_type="mixed tissues" /lab_host="SOLR" /note="Vector: plusscript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Combination of untreated and phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes / tubers, or roots."
BASE COUNT	216 a 168 c 170 g 215 t
ORIGIN	
Query Match	14.6% Score 350.6; DB 14; Length 769;
Best Local Similarity	75.2%; Pred. NO.2.6e-75;
Matches 464; Conservative	0; Mismatches 149; Indels 4; Gaps 2;
OY 674	ACTCTTGAAGATGGCTTTGTTGCTTCGTCCCGAAGTTTGATGACGTGAGGCT 733
Dd 87	ATTGTTAAATATAGCAATAGTAGGCTCTGCCTCGAAGATTGATGACGTGAGGTT 146
OY 734	ACCTGTGACTGAGACGCCAAATGACGGGCTTGTGTGATACAATGACAGGTTTTACG 793
Dd 147	ATCTGTTCTGGAGGGCCAATGCTGGTATTTGTACTAGACGAATAATCCTGTTTTATG 206
OY 794	CGATTGTCAGACCATCAACGAAAGTCAAGCCGTAAGAATTGGGCAATGAGAAC 853
Dd 207	CTATTGTGAAGCCACTTTATGACGAATTAGACCTGAAGTTGGGCATGGGGTTGACAG 266
OY 854	ATGTCAAATTAACATCACACCAGCTCTCGAGGAAAGCATGTATAACTGTCACTGAATC 913
Dd 267	ATGTAAATTTACCTTCTCCAGATGGCAAGCAAAACATGTATMACTGTCTCTTAAC 326
OY 914	ATTGACGCTTCAGTCGTCGTCTGCAAGTGAATTCGAAGAAACCCCTTTAGACGTGGCA 973
Dd 327	ACCTTAAGCTTCAGCTCTTTCTTCOAAGTGCATCAAGAAACCCCTTTTAGACATCGAG 386
OY 974	TACAGTATGCTATAGCTCTGCTCATTTTGGCAACCGAAGG--ACAAGATCATTTGC 1030
Dd 387	TGCAGTATGCTGTCGAGGGTCTGCAATACCTATCTTGGTTGCAATAGATTCATTTGCT 446
OY 1031	ACAAACTTTATGCAAGCTCTGATATAACAATTTAGGCTCCAAATACCTTTTACTCAT 1090
Dd 447	AGAGACTACTCTTCGAGGGTCTGCAATACCTATCTTGGTTGCAATAGATTCATTTGCT 506
OY 1091	ATCGGAACCAATGATACGCTGGGCTTCATTGATGACACCAAGATCGGTGATACCTTTG 1150
Dd 507	ATCGGAATCAGATTGACAGCAGTGGGCTCCCTCTTACACCTAAGCATTTGGCTTCCCTTC 566
OY 1151	CACCGTTTGCAATCAATCAATTCATGCTGGGAGTCAAAATGTCCTAATTTCCAGGCTG 1210

Db 567 CACCTTTTCCTCGATCACTTTAATGCAGAGAAATCAACTGAGAAAATTGCAAGCCTG 626  
OY 1211 AAGTAGCAAAACCTGG-CTTAGGTCTCTGAGCAATGACAACAGCTGTGTTGAGCT 1269  
Db 627 AAGTTGCAAGAGACTGGAATTAGGGTCGTCAGCAGCTATGACAAGTGCAGTTGTGCGGCC 686  
OY 1270 CTGTTACATTATCTTGG 1286  
Db 687 TTGCTTCATTATCTTGG 703

Search completed: May 3, 2003, 19:54:21  
Job time : 2265 secs

THIS PAGE BLANK (USPTO)



Result No.	Score	Query Match	Length	DB	ID	Description
1	229.2	9.6	557	6	US-09-909-745A-21	Sequence 21, Appl
2	194.4	8.1	577	11	US-60-288-292-33093	Sequence 33093, A
3	171.8	7.2	379	8	US-10-389-048-18322	Sequence 18322, A
4	158.2	6.6	675	7	US-09-837-604A-65010	Sequence 65010, A
5	150.4	6.3	539	6	US-09-909-745A-19	Sequence 19, Appl
6	140	5.8	388	7	US-09-837-604A-34763	Sequence 34763, A
c_7	135.8	5.7	585	7	US-09-837-604A-38723	Sequence 38723, A
8	113.8	4.7	296	6	US-09-615-606A-54496	Sequence 54496, A
c_9	107.2	4.5	249	6	US-09-909-745A-15	Sequence 15, Appl
10	103.2	4.5	249	6	US-09-909-745A-17	Sequence 17, Appl
11	73.6	3.1	1269	8	US-10-369-493-25849	Sequence 25849, A
12	57.6	2.4	1356	8	US-10-369-493-25596	Sequence 25596, A
13	43.8	1.8	1506	1	PCr-US02-40225-2218	Sequence 2218, Ap
14	43.8	1.8	1506	9	US-10-320-797-2218	Sequence 2218, Ap
c_15	42	1.8	2556324	7	US-09-947-911-95	Sequence 95, Appl
c_16	41.8	1.7	1664976	6	US-09-692-570-1	Sequence 1, Appl
17	41.4	1.7	1816	1	PCr-US02-40225-1218	Sequence 1218, Appl
18	41.4	1.7	1816	9	US-10-320-797-1218	Sequence 1218, Ap
19	41.4	1.7	3156	1	PCr-US02-40225-218	Sequence 218, Appl
20	41.4	1.7	3156	9	US-10-320-797-218	Sequence 218, Appl
c_21	41	1.7	360	7	US-09-837-604A-18934	Sequence 18934, A
22	40.4	1.7	1206	8	US-10-369-493-27427	Sequence 27427, A

23	40.4	1.7	3468	9	US-10-282-122A-10932	Sequence 10932, A
24	39.6	1.7	1536	8	US-10-369-493-46489	Sequence 46489, A
25	39.4	1.6	4455	8	US-10-391-339-5507	Sequence 5507, Ap
26	38.8	1.6	648	7	US-09-675-784A-5811	Sequence 5811, Ap
27	38.8	1.6	1141	6	US-09-806-708B-22	Sequence 22, Appl
28	38.4	1.6	247	7	US-09-837-604A-58235	Sequence 58235, A
29	38.4	1.6	527	7	US-09-837-604A-5480	Sequence 5480, Ap
30	38	1.6	405	7	US-09-837-604A-23709	Sequence 23709, A
31	37.8	1.6	264665	7	US-09-949-016-13747	Sequence 13747, A
32	37.6	1.6	333	7	US-09-837-604A-11362	Sequence 11362, A
33	37.4	1.6	9652	9	US-10-311-455-8892	Sequence 882, App
34	37.4	1.6	135050	7	US-09-949-016-18496	Sequence 14866, A
35	37.4	1.6	3673778	9	US-10-312-841-2	Sequence 2, Appl
36	37.2	1.6	289	7	US-09-837-604A-464	Sequence 464, Appl
37	37.2	1.6	2209	1	PCT-US02-41115-94	Sequence 94, Appl
38	37.2	1.6	17211	9	US-10-311-455-628	Sequence 628, App
39	37.2	1.6	50563	7	US-09-949-016-15821	Sequence 15821, A
40	37	1.5	411	6	US-09-615-606A-46691	Sequence 46691, A
41	37	1.5	540438	7	US-09-947-9011-322	Sequence 322, App
42	36.8	1.5	274	7	US-09-837-604A-10493	Sequence 10493, A
43	36.8	1.5	561	7	US-09-837-604A-64175	Sequence 64175, A
44	36.8	1.5	606	7	US-09-949-016-166383	Sequence 166383, A
45	36.8	1.5	6264	9	US-10-240-452-67	Sequence 67, Appl

## ALIGNMENTS

```

1 RESULT 1
2 US-09-909-745A-21
3 : Sequence 21, Application US/09909745A
4 : GENERAL INFORMATION:
5 :
6 : APPLICANT: Falco, S. Carl
7 :
8 : TITLE OF INVENTION: Squalene Synthesis Enzymes
9 :
10 : FILE REFERENCE: Bb112 US CIP
11 :
12 : CURRENT APPLICATION NUMBER: US/09/909,745A
13 :
14 : CURRENT FILING DATE: 2001-07-20
15 :
16 : PRIOR APPLICATION NUMBER: 09/433,242
17 :
18 : PRIOR FILING DATE: 1999-11-04
19 :
20 : PRIOR APPLICATION NUMBER: 60/107,241
21 :
22 : PRIOR FILING DATE: 1998-11-05
23 :
24 : NUMBER OF SEQ ID NOS: 26
25 :
26 : SOFTWARE: Microsoft Office 97
27 :
28 : SEQ ID NO 21
29 :
30 : LENGTH: 757
31 :
32 : TYPE: DNA
33 :
34 : ORGANISM: Oryza sativa
35 :
36 : US-09-909-745A-21

```

	Query Match	Similarity	Score	DB	Length
	Best Local	Similarity	Score	No. 2, 3e-52	757
	Matches	369; Conservative	0; Mismatches	233; Indels	0; Gaps
OY	1601	GAGTGTGTGATCTTCACACCATCAATGGTATGCTCAATAAAGAGTGGCAAAATTCG	1660		
Db	9	GATATGGGAGATCAATCTCCATCAATGGTATGCTTGAAGCAAGTGGCAAAATTCAG	68		
OY	1661	ATCCAGAGAGCAGAGAGAAATGGCAAAATTTGTCAATATGCAGAAATTTAAGACTGGAA	1720		
Db	69	ACCCACAGAAATCCAAAGAGAGACATGAGTAAATTTGGGGAGTTGCTAATTCAGTCTTGAGA	128		
OY	1721	CTAAGCTTAAGCATCTGAGCAAAATTAAGCTAAAGACCCTGGATGTTTATCTACAGTCA	1780		
Db	129	ACCAACTGAGGAACCTAAACAAACTGTGCAAGATACAGGGAGGCTATGATTCGTTT	188		
OY	1781	TTAAGTCTTGTATGTGCTTACTCTCGAAGAAAGGGGTGTTAATCATCTGAACCAATCA	1840		
Db	189	TACGATCTCTGTGTGCTCTCACAGTCAGTAAGTGAACAGAGGGCTACCAATCAACATC	248		
OY	1841	ACGAAGCCATTATTAAGACTCTTAAGAGGCAAGAGACCTATGTTGAGATCAGAATTC	1900		
Db	249	AAGAACTAATGTTTGATCATATTACATGGCCGCAAGAGATCTTCTCTGAAATTAAGCTTC	308		



;; CURRENT APPLICATION NUMBER: US/09/837, 604A  
;; CURRENT FILING DATE: 2001-04-18  
;; PRIOR APPLICATION NUMBER: US 60/197, 872  
;; PRIOR FILING DATE: 2000-04-19  
;; NUMBER OF SEQ ID NOS: 81288  
;; SEQ ID NO 65010  
;; LENGTH: 675  
;; TYPE: DNA  
;; ORGANISM: Oryza sativa nipponbare  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: LIB3598-023-P1-K6-D6  
US-09-837-604A-65010

Query Match 6.6%; Score 158.2; DB 7; Length 675;  
Best Local Similarity 60.3%; Pred. No. 7.4e-33;  
Matches 279; Conservative 0; Mismatches 183; Indels 1; Gaps 1;

Qy 1741 AAATTAGCTAAAGACCACTGGGATGTTATCTACGAGTCATTAGCTTTAGTGTCTT 1800  
Db 1 AACTTGTCTAGACACCTGGGAGCCCTATGTAATCTGTTTACGATCCGTAGTGTCTC 60  
Qy 1801 AC-TTCTGAAAAGTGGGTGTTACATGCTACTGTAACCAATCAAGAGCATTTATAAGA 1859  
Db 61 ACCTCAGTAAGGTGACAGAGAGTGGCTACCAATCAATCAAGAGCATTTATAAGTATC 120  
Qy 1860 ACTTTAGAGCAGAGAGAGATGTTAGAGATTCAGAAATTTTATGCGTACAGATGGTGA 1919  
Db 121 ATTACTGCGCCGAGAGAGATGTTCTTCTTGAATAATAGGCTCTATGCGAGATGGCAT 180  
Qy 1920 GCGGCTAGCGTCCGATAGAGCTGAATCTCAAACTCACTTTTGGATTCTACATAG 1979  
Db 181 AGCAGCTGTGTTCCAAATGAGCCAGAAATCAAACTCAACTTTCAGTCCACTATGAA 240  
Qy 1980 TGCTGAGAGATTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2039  
Db 241 TATGAGGGGTGTTCTACTAGCTGAGTTCCTGGGCGCGTGGCTTGAATGAGTTCG 300  
Qy 2040 AATCACTTTAGGAGATTCGCGACCAAACTGACCCAGCATGTCGACATGTTT 2099  
Db 301 AGTATTTTGGGTGAGAGAGATGCTGTAGCCAAAGCTTGGAGCTGAGTGTCT 360  
Qy 2100 GCGCTTTGTTGAGAGAGATCCACATGCGCTTTCCTAGAAAGTGTGATCCAGAAC 2159  
Db 361 CCCCTTTCTTGTTCGAGAGATCCCGGAGGTGTTCATTTGGAAGGTGGACCAAGAAC 420  
Qy 2160 CACATGTTACTTACGAGCGTTTCTCATTTACCTTGAGTAA 2202  
Db 421 AAGGAGGTGTCACCGCTGTATCATTCATCAATAAATGAA 463

## RESULT 5

US-09-909-745A-19  
;; Sequence 19, Application US/09909745A  
;; GENERAL INFORMATION:  
;; APPLICANT: Falco, S. Carl  
;; APPLICANT: Farnodu, Omolayo O.  
;; TITLE OF INVENTION: Squalene Synthesis Enzymes  
;; FILE REFERENCE: B1112 US CIP  
;; CURRENT APPLICATION NUMBER: US/09/909, 745A  
;; CURRENT FILING DATE: 2001-07-20  
;; PRIOR APPLICATION NUMBER: 09/433, 242  
;; PRIOR FILING DATE: 1999-11-04  
;; PRIOR APPLICATION NUMBER: 60/107, 241  
;; PRIOR FILING DATE: 1998-11-05  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: Microsoft Office 97  
;; SEQ ID NO 19  
;; LENGTH: 539  
;; TYPE: DNA  
;; ORGANISM: Oryza sativa  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (287)

;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (440)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (454)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (460)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (465)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (486)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (488)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (495)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (530)  
;; OTHER INFORMATION: n=A, C, G, or T  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (533)  
;; OTHER INFORMATION: n=A, C, G, or T  
US-09-909-745A-19

Query Match 6.3%; Score 150.4; DB 6; Length 539;  
Best Local Similarity 62.6%; Pred. No. 9.6e-31;  
Matches 266; Conservative 0; Mismatches 157; Indels 2; Gaps 2;

Qy 1601 GAAGTGTGATGCTCCACACCATCAATGATGTGAGTAAAGAGTGAATGTCTG 1660  
Db 2 GTATGGAGAGATCATCCATCATATGTTGATCTGTGAACAGTGGCAGAAAGTCTG 61  
Qy 1661 ATCCAGAGAGGACGAGAGAAATGCGAGAAATTTGTACAGATCCAAATTTAGAACTGAAA 1720  
Db 62 ACCCTCAGAAATCCAGAGAGATGAGTAATTTGGGATTCCTAATTCAGTGTGAGA 121  
Qy 1721 CTAAGCTAAAGATCTGACCAATTTAGCTAAAGACACATGCGATTTATCTACAGTCA 1780  
Db 122 ACCAAGCTGAGAACTTAACAACTTGTGAAGATCAGGAGAACCTATGAAATCTGTTT 181  
Qy 1781 TTAAGCTGTGATGCTTACTTCTGAAAAGTGGGTGTTACATGCTACTGAAACCAATCA 1840  
Db 182 TAGCATCTGTAGTGTCTACAGTCAATGAGAGAGTGGCTACCAATCAACATC 241  
Qy 1841 ACGAAGCATTATTAAGAACTCTTAGAGGCAAGAACTATGTTGAGAGTCAAGAAATTC 1900  
Db 242 AAGAACTAATTTAGATCATTTAGTGGCGGCAAGAGATGCTTCNTGAATAAGAGCTTC 301  
Qy 1901 TTATG-CGTACAGATGGGTGAGGCGGCTAGCTTCGATAGAGCTGAAATCTCAA 1959  
Db 302 AATATCCAAAGATGGGATAGCATGCTGTGTTCAATTTGACCAAGAAATCAAACTCAA 361  
Qy 1960 CTTTGGATTTAGCAATGAGTGTCTGAGAGAGTGTCTACTGCTGTTGCTGAGAGCTGCT 2019  
Db 362 CTTCTGATGCCATATGATATGAGAGGCTTTCTACTACTGAG-ATTCTTGAGGCGGCT 420  
Qy 2020 GGATT 2024  
11 11









THIS PAGE BLANK (uspto)



GenCore version 5.1.4.p5.4578.  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 18:09:42 ; Search time 3519. Seconds

(without alignments) : 17118.898 Million cell updates/sec

Title: US-09-988-863a-1  
Perfect score: 2396  
Sequence: 1 gtcgaccacgcgtccggc.....ttctcaaaaaaaaaaaaaa 2396

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/US06000.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US06001.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US07000.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US08000.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US08100.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US08200.COMB.seq:\*  
7: /cgn2\_6/ptodata/1/pna/US08300.COMB.seq:\*  
8: /cgn2\_6/ptodata/1/pna/US08400.COMB.seq:\*  
9: /cgn2\_6/ptodata/1/pna/US08500.COMB.seq:\*  
10: /cgn2\_6/ptodata/1/pna/US08600.COMB.seq:\*  
11: /cgn2\_6/ptodata/1/pna/US08700.COMB.seq:\*  
12: /cgn2\_6/ptodata/1/pna/US08800.COMB.seq:\*  
13: /cgn2\_6/ptodata/1/pna/US08900.COMB.seq:\*  
14: /cgn2\_6/ptodata/1/pna/US09000.COMB.seq:\*  
15: /cgn2\_6/ptodata/1/pna/US09100.COMB.seq:\*  
16: /cgn2\_6/ptodata/1/pna/US09200.COMB.seq:\*  
17: /cgn2\_6/ptodata/1/pna/US09300.COMB.seq:\*  
18: /cgn2\_6/ptodata/1/pna/US09400.COMB.seq:\*  
19: /cgn2\_6/ptodata/1/pna/US09500.COMB.seq:\*  
20: /cgn2\_6/ptodata/1/pna/US09600.COMB.seq:\*  
21: /cgn2\_6/ptodata/1/pna/US09700.COMB.seq:\*  
22: /cgn2\_6/ptodata/1/pna/US09800.COMB.seq:\*  
23: /cgn2\_6/ptodata/1/pna/US09900.COMB.seq:\*  
24: /cgn2\_6/ptodata/1/pna/US10000.COMB.seq:\*  
25: /cgn2\_6/ptodata/1/pna/US10100.COMB.seq:\*  
26: /cgn2\_6/ptodata/1/pna/US10200.COMB.seq:\*  
27: /cgn2\_6/ptodata/1/pna/US10300.COMB.seq:\*  
28: /cgn2\_6/ptodata/1/pna/US10400.COMB.seq:\*  
29: /cgn2\_6/ptodata/1/pna/US10500.COMB.seq:\*  
30: /cgn2\_6/ptodata/1/pna/US10600.COMB.seq:\*  
31: /cgn2\_6/ptodata/1/pna/US10700.COMB.seq:\*  
32: /cgn2\_6/ptodata/1/pna/US10800.COMB.seq:\*  
33: /cgn2\_6/ptodata/1/pna/US10900.COMB.seq:\*  
34: /cgn2\_6/ptodata/1/pna/US11000.COMB.seq:\*  
35: /cgn2\_6/ptodata/1/pna/US11100.COMB.seq:\*  
36: /cgn2\_6/ptodata/1/pna/US11200.COMB.seq:\*  
37: /cgn2\_6/ptodata/1/pna/US11300.COMB.seq:\*  
38: /cgn2\_6/ptodata/1/pna/US11400.COMB.seq:\*  
39: /cgn2\_6/ptodata/1/pna/US11500.COMB.seq:\*  
40: /cgn2\_6/ptodata/1/pna/US11600.COMB.seq:\*  
41: /cgn2\_6/ptodata/1/pna/US11700.COMB.seq:\*  
42: /cgn2\_6/ptodata/1/pna/US11800.COMB.seq:\*  
43: /cgn2\_6/ptodata/1/pna/US11900.COMB.seq:\*

Result No.	Score	Query Match	Length	ID	Description
1	2396	100.0	2396	US-09-988-863a-1	Sequence 1, Appl1
2	840.2	35.1	1509	PCT-US02-24048A-5	Sequence 5, Appl1
3	840.2	35.1	1509	US-10-036-959B-5	Sequence 1155, Ap
4	759	31.7	119420	US-09-803-736-1155	Sequence 121976,
5	569.6	23.8	576	US-09-654-617-121976	Sequence 9087, Ap
6	569.6	23.8	576	US-09-684-016-121976	Sequence 6586, A
7	518.4	21.6	1956	US-10-219-999-9087	Sequence 2752, Ap
8	486.6	20.3	1892	US-09-708-427-6586	Sequence 13678, A
9	473.4	19.8	475	US-09-649-165A-2752	Sequence 5570, Ap
10	463.6	19.3	1889	US-09-708-427-83406	Sequence 154090,
11	454.4	19.0	1887	US-09-708-427-66288	Sequence 154090,
12	443	18.5	443	US-09-638-111-871	Sequence 871, App
13	443	18.5	443	US-09-770-444-851	Sequence 851, App
14	425.4	17.8	728	US-09-654-617-192867	Sequence 192867,
15	425.4	17.8	728	US-09-684-016-192867	Sequence 192867,
16	425.4	17.8	728	US-09-988-863a-4	Sequence 4, Appl1
17	421.4	17.6	458	US-09-333-534-13646	Sequence 13646, A
18	418.2	17.5	511	US-09-333-534-13678	Sequence 13678, A
19	411.2	17.2	457	US-09-333-534-5570	Sequence 5570, Ap
20	402.2	16.8	448	US-09-654-617-154090	Sequence 154090,
21	402.2	16.8	448	US-09-684-016-154090	Sequence 154090,

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

```
C 22 400.4 16.7 840 19 US-09-505-532-46021 Sequence 46021, A
C 23 400.4 16.7 840 31 US-09-819-091A-46021 Sequence 46021, A
C 24 390 16.3 501 17 US-09-333-534-2389 Sequence 2389, A
C 25 383.2 16.0 489 25 US-09-654-617-134754 Sequence 134754, A
C 26 383.2 16.0 489 25 US-09-684-016-134754 Sequence 134754, A
C 27 382.8 16.0 1470 25 US-09-654-617-259014 Sequence 259014, A
C 28 382.8 16.0 1470 27 US-09-684-016-259014 Sequence 259014, A
C 29 381 15.9 417 17 US-09-333-534-14375 Sequence 14375, A
C 30 362.4 15.1 364 25 US-09-654-617-181171 Sequence 181171, A
C 31 362.4 15.1 364 25 US-09-684-016-181171 Sequence 181171, A
C 32 358.8 15.0 421 17 US-09-333-534-16350 Sequence 16350, A
C 33 329.2 13.7 680 28 US-09-705-926-13137 Sequence 13137, A
C 34 307 12.8 1665 42 US-10-219-999-23255 Sequence 23255, A
C 35 302 12.6 1053 42 US-10-219-999-10586 Sequence 10586, A
C 36 302 12.6 1053 75 US-09-312-544-858 Sequence 858, A
C 37 299.2 12.5 611 37 US-09-988-863A-3 Sequence 3, A
C 38 286 11.9 460 18 US-09-404-520-9784 Sequence 9784, A
C 39 286 11.9 460 57 US-60-138-103-9068 Sequence 9068, A
C 40 247.8 10.3 519 28 US-09-705-926-20673 Sequence 20673, A
C 41 242.6 10.1 517 28 US-09-705-926-20669 Sequence 20669, A
C 42 229.2 9.6 757 18 US-09-433-242-21 Sequence 21, A
C 43 229.2 9.6 757 18 US-09-433-242-21 Sequence 21, A
C 44 229.2 9.6 757 34 US-09-909-745-21 Sequence 21, A
C 45 227 9.5 40548 20 US-09-534-859-716 Sequence 716, A
```

## ALIGNMENTS

```
RESULT 1
US-09-988-863a-1
; Sequence 1, Application US/0988863A
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Plant phosphomevalonate kinases
; FILE REFERENCE: Le A 35 018
; CURRENT APPLICATION NUMBER: US/09/988, 863A
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2396
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (685)..(2199)
US-09-988-863a-1
```

```
Query Match 100.0%; Score 2396; DB 37; Length 2396;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GTGACCCAGCGTCGCGGCGACCTTCTCTTAAAGCAACACATATAGTAG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 GTGACCCAGCGTCGCGGCGACCTTCTCTTAAAGCAACACATATAGTAG 60
OY 61 AAGCAAACTGGGGAAGATGAGATGAGTGAAGCAAAACCGTATACCGTTGCT 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 AAGCAAACTGGGGAAGATGAGATGAGTGAAGCAAAACCGTATACCGTTGCT 120
OY 121 TCAGAGTGGCGAAGCAACCGACCGCTTAAACCAAAATTCCTCAAAAGAAATTCGCGATCG 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 TCAGAGTGGCGAAGCAACCGACCGCTTAAACCAAAATTCCTCAAAAGAAATTCGCGATCG 180
OY 181 GTTGTCTACTGTTCAAAACCTGGTGGCGAAGCAAGAACTGGTGGTGGTGGTGGTGG 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 GTTGTCTACTGTTCAAAACCTGGTGGCGAAGCAAGAACTGGTGGTGGTGGTGGTGG 240
OY 241 GTTGTCTACTGTTCAAGATCCAGAGCTAGTTGGTTTATTTTCAAGATTTTCG 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 GTTGTCTACTGTTCAAGATCCAGAGCTAGTTGGTTTATTTTCAAGATTTTCG 300
```

```
OY 301 TTCTCTTTTCATCGGCGACGACGATGAGTTTCTGTCAAAACGTTAAGCATCCAGCTC 360
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 TTCTCTTTTCATCGGCGACGACGATGAGTTTCTGTCAAAACGTTAAGCATCCAGCTC 360
OY 361 GAGGTGACAGTAAAGAAAGAAACACAGCTTTGTGTAGATGACGGGAAAGCTGTCT 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 361 GAGGTGACAGTAAAGAAAGAAACACAGCTTTGTGTAGATGACGGGAAAGCTGTCT 420
OY 421 CGATCCGCTTCGATCGACGAGGAAATACGTTTCGATCCGTTTCGATCCAAATCGAGAGT 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 CGATCCGCTTCGATCGACGAGGAAATACGTTTCGATCCGTTTCGATCCAAATCGAGAGT 480
OY 481 TTGAGGATCTAAATCGAAATTCGATTAATCATCTCAATCTCTTGAAGATCCG 540
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 481 TTGAGGATCTAAATCGAAATTCGATTAATCATCTCAATCTCTTGAAGATCCG 540
OY 541 AATCCGATCTAACCCACTCTCTACCGCGGATTTCTGCGCGGCAATTCGAATTA 600
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 AATCCGATCTAACCCACTCTCTACCGCGGATTTCTGCGCGGCAATTCGAATTA 600
OY 601 TCCGATATTTCCGCGCATATCCAAATCGACAGCTGAGTGAATCGGGTTTGTATCAGC 660
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 601 TCCGATATTTCCGCGCATATCCAAATCGACAGCTGAGTGAATCGGGTTTGTATCAGC 660
OY 661 GATTATCTTTGTACCTCTTTGAAGATGGCTGTCTCTCTCTCGGAAAGTTTG 720
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 661 GATTATCTTTGTACCTCTTTGAAGATGGCTGTCTCTCTCTCGGAAAGTTTG 720
OY 721 ATGACTGAGAGCTACCTTGTACTGAGAGCAAGCAATCGAGGCTTGTGAGTAAAT 780
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 721 ATGACTGAGAGCTACCTTGTACTGAGAGCAAGCAATCGAGGCTTGTGAGTAAAT 780
OY 781 GCACGGTTTTACGGATTTGAAAGCCAAATCAACGAAAGTCAAGCCTGAAATTTGGCA 840
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 781 GCACGGTTTTACGGATTTGAAAGCCAAATCAACGAAAGTCAAGCCTGAAATTTGGCA 840
OY 841 TGGAAATGACAGATGTCAAATTAACATCAACAGCTCTCGAGAAAGATCTATATAA 900
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 841 TGGAAATGACAGATGTCAAATTAACATCAACAGCTCTCGAGAAAGATCTATATAA 900
OY 901 CTGTCACTGATCATTTGACCTCTCACTGCTGCTGCTCAAGTGAATTAAGAAACCCCTTT 960
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 901 CTGTCACTGATCATTTGACCTCTCACTGCTGCTGCTCAAGTGAATTAAGAAACCCCTTT 960
OY 961 GTAGAGCATGAGATACAGTATGCTATGATGCTGCTCATTGTCGAACCGAAGAGCAAA 1020
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 961 GTAGAGCATGAGATACAGTATGCTATGATGCTGCTCATTGTCGAACCGAAGAGCAAA 1020
OY 1021 GAATCATTTGCAAAACCTTATTTGCAAGGCTTGTGATTAACATATTAAGCTCCATAGAC 1080
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1021 GAATCATTTGCAAAACCTTATTTGCAAGGCTTGTGATTAACATATTAAGCTCCATAGAC 1080
OY 1081 TTTTACATATGGAACAGATGAGATGAGTGGGCTTCATTCAGACAGAAATCGCTG 1140
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1081 TTTTACATATGGAACAGATGAGATGAGTGGGCTTCATTCAGACAGAAATCGCTG 1140
OY 1141 GGTACCTTCGACCGTTTGCATCAATCAATCTCGGAGTCAAAATGCTGTAAT 1200
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1141 GGTACCTTCGACCGTTTGCATCAATCAATCTCGGAGTCAAAATGCTGTAAT 1200
OY 1201 TCCAGAGCTGAGTGAAGAAACAGGCTTGTGCTCTGACGAATGACAAGAGCTGTG 1260
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1201 TCCAGAGCTGAGTGAAGAAACAGGCTTGTGCTCTGACGAATGACAAGAGCTGTG 1260
OY 1261 GTTGCACCTCTGTTTACATTAATCTTGAGTGGTGAACCTATCTGATCTAATGAAGAAGA 1320
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1261 GTTGCACCTCTGTTTACATTAATCTTGAGTGGTGAACCTATCTGATCTAATGAAGAAGA 1320
OY 1321 AAGTTGGCTGTTCTGATCTAGATGTTATCCATATGATACAAACGTCATCTGCTT 1380
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1321 AAGTTGGCTGTTCTGATCTAGATGTTATCCATATGATACAAACGTCATCTGCTT 1380
OY 1381 GCACAGGGAAGGTGGAAGTGGGTTTGATGTACCTGTCTGTATGGAAGTCAAGCT 1440
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1381 GCACAGGGAAGGTGGAAGTGGGTTTGATGTACCTGTCTGTATGGAAGTCAAGCT 1440
```



Db	775	GTGCTTCCTCCTGCTCAGGATGCTGGGAAAGAAATTCATTCACGAGAACTCATTTCTTAC	834
Qy	1522	ATTTTGAAGGGAATAATGGGCAATTAAGAGAACTGAGTTCTTTTACCACTACTATGAAT	1581
Db	835	ATCCTAAAGGAAATAATGGGACCAATGAGAGACATATCTTTTCCCTTGCCACATTTGATGAGC	894
Qy	1582	CTTTTCCCTGGAGAACTGGAGAGGGGATGCTCCACACATTCATGCTGATGAGTACAGA	1641
Db	895	CTTGCTACTAGGTGAGCCAGGAACTGGAGGATCTTCCACGCTATCAATGGTATGACTCTCA	954
Qy	1642	AAGAAGTGGCAAAATGTTGTGATCCAGAGAAAGGACGAGAAAACTGGCAGAAATTTGCAGAT	1701
Db	955	AAGAAATGGCAGGAAGTGTGATCTACGAAATCCCAAGAAACATGGAGAAAGTTGTACAGAG	1014
Qy	1702	GCAAAATTTAGACTGGAAACTTAAGCTTAACGATCTGAGCAAAATTTAGCTTAAGACCACTGG	1761
Db	1015	GCAAAATTCAGCACTTGGAAAGCAATTCAAATTTTAAAGACGCTCCAGAAACATATTGG	1074
Qy	1762	GATGTTTATCTAGACGATCTTAAGTCTGTGAGTGTCCTTACCTTCGAAAGAGGGCTGTTA	1821
Db	1075	GAGCGTGTAAATATGTGATGAGACAGTTGCAGACAAAAACTCAGAGAAAGTGATGAG	1134
Qy	1822	CATGCTACTGACCAATCAACGAGCCATTTTAAAGAACTCTTTAGAGCAGAGAAAGCT	1881
Db	1135	CAGGCACTGAGACCCAGCCAGGAAGCAAGCTTTGTTAAAGCATTTATTTAGATCAAGAAATGCC	1194
Qy	1882	ATGTTGAGGATCAGAAATTTCTTATGCTGAGATGGTGGAGGGGCTAGGCTTCCGATAGAG	1941
Db	1195	ATGCTTACGATCAGAAATTTACATGCGCCAGCAATGGTGAGGCTCAGAGTGTCCGATAGAG	1254
Qy	1942	CCTAAATCTCAAACCTCAACTTTTGGATCTTACAAATGATGCTGAGAGAGTTCTACTTGGT	2001
Db	1255	CCTGAATCCAGACTGCACCTTTTGGATACTACTATGATATGATGATGAGTCTTTGTTGGCT	1314
Qy	2002	GGTGTCTCTGAGAGCTGTGTGATTTTGATGCCATATTTGCACATCTTTAGGGATTCGCCG	2061
Db	1315	GGAGTTCCTGGAGAGAGTGGGTTTGATGACAGCTTGTGCTTTTACCTTTAAGGAGACTCTGCT	1374
Qy	2062	ACCAAACTGACCCAGGAGTGGAGTTGGCCAAATGTTTGGCCCTTGTGTGTGAGAGAGAT	2121
Db	1375	ACCAATATGGGCAAAAGCTTGGAGTTCACCTAATGTTCTGGCCCTGTTGGTTGAGAGAAGC	1433
Qy	2122	CCACATGGCGTTTCCCTAGAAAGTGGTATCCACGAAACCACTATGTTTACTTCAGCGCTT	2181
Db	1435	CCATATGCTGTTTGTGTAGAAAGCGGAGATCCAAAGAACCAAGGAATTCACAAACAGCTGTT	1494
Qy	2182	TCATCAATTCACCTT 2196	
Db	1495	TTTGGAGTTCATATT 1509	

### RESULT 3

US-10-036-959B-5

; Sequence 5, Application US/10036959B

GENERAL INFORMATION:

APPLICANT: E. I. du Pont de Nemours & Company

APPLICANT: Hallahan, David L.

APPLICANT: Keiper-Hrynko, Natalie

1 TITLE OF INVENTION: Genes Involved in the Biosynthesis of Isopentenyl Diphosphate in

;; TITLE OF INVENTION: brasiliensis Latex

FILE REFERENCE: CL-1792

FILE REFERENCE: CU 1/52  
CURRENT APPLICATION NUMBER: US/1

CURRENT FILING DATE: 2002-05-10

PRIOR APPLICATION NUMBER: 60/307-637

PRIOR FILING DATE: 2001-07-

NUMBER OF SEQ ID NOS: 16

NUMBER OF SEQ ID NOS: 16

```

; SOFTWARE: MIC
; CEO ID NO 5

```

; SEQ ID NO 5

LENGTH: 1509  
TYPE: DNA

```

; TYPE: DNA

```

Query Match	35.1%;	Score 840.2;	DB 38;	Length 1509;
Best Local Similarity	73.1%;	Pred. No. 1.9e-207;		
Matches 1108; Conservative	0;	Mismatches 398;	Indels 9;	Gaps 2;

OY	685	ATGCTGTTGGTGGTCTGTCGCTCGGGAAGTTTATGACATGAGGCACTACTGTATGAC	744
Dd	1	ATGGCTAGTGTGCTTCTGCTCCGGGTAAGGTGTATGACATGAGGCTTACTCATTTGG	60
OY	745	GAGAAAGCCAAATGCAAGGGCTGTGTGTGATGACAAATGCAAGCGGTTTTACGCGATTGTGAAG	804
Dd	61	GAAAGACCCCAATGCAAGGATTTGTACTGACGACAAATGCTGATTCATGCTCTGTGTGAAG	120
OY	805	CCATTCACGAAGAAGTCAAGCCTGAAAGTTGGGCACTGGAATATGACAGATGCTCAATTA	864
Dd	121	CCTATTATTACGTAATCAACCTGATAGTGGGCAATGGCATGGCATGTATGTAATTA	180
OY	865	ACATCACACAGCCTCGAGAGAAAGCATATATAACTGTCACTGATTCATTTGCTCTT	924
Dd	181	ACATCTOCCCAACTGACGAAGGAAGCTGTACAAATGTGCACTGAAAAATTTACTCTGT	240
OY	925	CAGTCTGTGTGCGAAGTGAATTCAGAAACCCCTTTGTAGAGCATGCGATACATGATGCT	984
Dd	241	CAGTGTGCTCTTCAAGTGTGATGAGAACCCATTTGTGTGAACCAAGCACTGCAATTTCTT	300
OY	985	ATAGTGTGCTCATTTTGGCAACCGAAGG---ACAAGAATCATGTGACCAAACTCTTA	1041
Dd	301	GTACGAGCTGCACATGCAACACTTGCACAAAGATATAGAAATGTCTTAACCAAGCTACT	360
OY	1042	TTGCAAGGCTTGTGATTAACATATTAAGGCTCCATGACTTTTACTCATATGGAACAG	1101
Dd	361	TTGCAAGGCTTGTGATTAACATATTAAGTATGATACCAATGACTTATTCATACCGAATAG	420
OY	1102	ATGATAGGCTGGGCTTCATATGACACCGAGATGCTGGATACCTTTCACCGTTTGTGA	1161
Dd	421	ATTGAGCATGTGGACTCCCTTTTGACCCAGAAATCAATTGGCTGCACTTCTTCTTTTCC	480
OY	1162	TCATCATCATTCATGCTGGCGAGTCAAAATGTCATTTCCAAAGCTGAAGTAGCAAAA	1221
Dd	481	TCATCATCATTCATGATGATGAGAGCAAAATGAGCAAAATCGCAAGGCTGAGTACTTAA	540
OY	1222	ACTGCTTAGGTTCTTCTGCAAGCAATGACAAAGCTGTGGTGTGACAGCTGTGTTACATAT	1281
Dd	541	ACTGATTTGGTTATATGAGAGCAATGACCAAGCTGTGTGCTGCTTACTTATCATC	600
OY	1282	CTTGAGAGTGTGACCTATGCTGATCCATGTAAGAAGAAAGTTTGGCTGTCTGATCTA	1341
Dd	601	CTTGATTTGGTTGATCTTTCATCTCTGTTAAGAAAGAAATTT-----TTGTCATCTT	654
OY	1342	GATGTTTATCATATGATGATGACAAACGCTCATTTGCTTTCACAAAGGAAGTGGCAAGT	1401
Dd	655	GATTTGTGATATATATAAGCCCAAACTGCCCATTTGATTTGCAAGGAAGAAAGTCCGAGT	714
OY	1402	GGGTTTATGTCACTGTGCTGCTATGAGAAAGCAGAGTATAGTGGCTTCTCTCAGAA	1461
Dd	715	GGATTTGATGTTACTTCTGCAAGTTTATGGAGTCAATGATAGTGGCTTCTCTCAGAA	774
OY	1462	GCTTTGTCATTTGCTGAGGTTTGCAGTAAACAGCTCCATTAATGAAATTTTGGTACA	1521
Dd	775	GTCGTTCTCTGCTGCAAGATGCGGGAAAGAAATTCATTTACAGGAATCATTTCTTAC	834
OY	1522	ATTTTGAAGGGAATGAGCAATTAAGAGACTGAGTTCTCTTTACCAACCACTGATGAT	1581
Dd	835	ATTCCTAAAAAGGAATATGGGACATATGAGAGACTATTTTTCCTTGGCACCATTTGTAGC	894
OY	1582	CTTTTCTGTGAAAGCTGGAAGGTGGATCTCTGACACATTCATAGGTTAGTGCAGTA	1641
Dd	895	CTGCTACTAGGTGAGCCAGGAATCTGGAGGATCTTCCACGCAATCAATGGTATAGTCTTA	954
OY	1642	AAGAAGTGGCAAAATGCTGATTCAGAGAAAGCAAGCAAAATGCGAGATTTGTACAGAT	1701
Dd	955	AAGAATATGCGAAGTCTGATACTCAGAAATCCCAAGAAACATGGAAGAAATGTTGCAGAG	1014
OY	1702	GCAATTTAGAACTGGAACCTAAGCTTAACGATCTGAGCAAAATTAAGTAAGCACTGCG	1761

```

Db 1015 GCAATTCAGACCTTGAACCAATTCATATTTTACAGAGCTCGCAGAAAGACATTGG 1074
Qy 1762 GATGTTTATCTACAGATATTAAGCTCTGTAGTGTCTACTTCTGAAAAGGCGTGTTA 1821
Db 1075 GACGGTAAATATGTATAGACAGTGTGACAGAAAAAATCTAGAGAGGATTTAG 1134
Qy 1822 CATGCTACTGAAACCAATGAAGAGCCATTTATTAAGAAGCTTTAGAGCAAGAGACCT 1881
Db 1135 CAGGCACTGAACCCAGCCGAGAGCAAGTGTTAAGCATTTATTAAGATCAAGAAATGCC 1194
Qy 1882 ATGTTGAGAGATGAGAAATCTTATGCGTCAGATGGGTGAGCGCGCTTCCGATAGAG 1941
Db 1195 ATGCTTACAGATCAGAAATTAACATGCGCAGATGGGTGAGCGCTTCCGATAGAG 1254
Qy 1942 CCTGAATCTCAAACTCAACTTTTGGATTTCTACATGAGTGGCTGAAGAGCTTCTACT 2001
Db 1255 CCTGAATCTCAAACTCAACTTTTGGATTTCTACATGAGTGGGTGAGCGCTTCTACT 1314
Qy 2002 GGTGTTCTGAGAGCTGAGATTTGATGCTTATTTGCAATCTTTAGGGGATTCGGC 2061
Db 1315 GGAGTCTCTGAGAGCAGGAGGCTTTGATGCTTCTGCTTACTTATAGGGAGCTCTGGT 1374
Qy 2062 ACCAAATGACCCAGAGCATGAGATTCGACAAATGTTTGGCTTGTGTGAGAGAGAT 2121
Db 1375 ACCAATGTGCAAAAGCTTGGATGACTCACTAATGTTTGGCTTGTGTGAGAGAGAC 1434
Qy 2122 CCACATGGCGTTGGCTGAAGAGTGTATGACAGAACCAATGATTTACTTACAGCGTT 2181
Db 1435 CCTAATGGTGTGTTGTATGAAGGCGGAGATCCAGAACCAAGAAATCAACACAGCTGT 1494
Qy 2182 TCATCAATTCACCTT 2196
Db 1495 TTTGAGTTCAATTT 1509

```

```

RESULT 4
US-09-803-736-1155/c
: Sequence 1155, Application US/09803736
: GENERAL INFORMATION:
: APPLICANT: Bush, David F.
: APPLICANT: Levin, Irena M.
: APPLICANT: Norris, Susan R.
: APPLICANT: Rounsley, Steven D.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: Plant polymorphic Markers and uses thereof
: FILE REFERENCE: 38-10115493D
: CURRENT APPLICATION NUMBER: US/09/803,736
: PRIOR FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/534,859
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983, 0206CPUS01 38-10
: NUMBER OF SEQ ID NOS: 1582
: SEQ ID NO 1155
: LENGTH: 119420
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-803-736-1155

```

```

Query Match 31.7%; Score 759; DB 31; Length 119420;
Best Local Similarity 72.6%; Pred. No. 1,76-185;
Matches 1283; Conservatively 0; Mismatches 0; Indels 484; Gaps 4;

```

```

Qy 1100 AGATAGATCGGCTGGCTTCCATTGACAGAAATCGCTGGGTACCTTGACCGCTTGG 1159
Db 29078 AGATAGATCGGCTGGCTTCCATTGACAGAAATCGCTGGGTACCTTGACCGCTTGG 29019
Qy 1160 CATCAATCAATCAATCTCGGAGTCAAAATGCTATTTCCAAAGCTGAAGTAGCAA 1219
Db 29018 CATCAATCAATCAATCTCGGAGTCAAAATGCTATTTCCAAAGCTGAAGTAGCAA 28959
Qy 1220 AAATGCGTTAGTCTTCTGACAGCAATGACAAAGCTGTGGTTCACAGCTCTGTACAT 1279

```

```

Db 28958 AAATGCGTTAGTCTTCTGACAGCAATGACAAAGCTGTGGTTCACAGCTCTGTACAT 28899
Qy 1280 ATCTTGGAGTGGTTTACCTATCTGATCATGTAAGAAGAAAGTTGGCTTGTGATC 1339
Db 28898 ATCTTGGAGTGGTTTACCTATCTGATCATGTAAGAAGAAAGTTGGCTTGTGATC 28839
Qy 1340 TAGATGTTATCCATATGATGATGACAAAGCTGCTATTTGCTTGCACAAAGAGTGGAA 1399
Db 28838 TAGATGTTATCCATATGATGATGACAAAGCTGCTATTTGCTTGCACAAAGAGTGGAA 28779
Qy 1400 GTGGGTTGATGATCAGTGTGCTGTCTATGAAAGTACAGCTTATGTTGCTTCTCCAG 1459
Db 28778 GTGGGTTGATGATCAGTGTGCTGTCTATGAAAGTACAGCTTATGTTGCTTCTCCAG 28719
Qy 1460 AAGTCTTGTCAATTCCTC----- 1477
Db 28718 AAGTCTTGTCAATTCCTCAGTTCAATACAGAAAGCTTGTATATTAATTCCTCAGTAA 28659
Qy 1478 -----AGTTGCAATACAGTGTGCCATTA 1504
Db 28658 AGGAAGGAATTTTACTGATCCCAAGTTTGTAGGTTGACAGTATGCTCCATTTAA 28599
Qy 1505 ATGAAGTATTTGTCATATTTTGAAGGAAATGGGACAAATGAAGAACTGATCTCTT 1564
Db 28598 ATGAAGTATTTGTCATATTTTGAAGGAAATGGGACAAATGAAGAACTGATCTCTT 28539
Qy 1565 TACACCACTGATGATCTT----- 1584
Db 28538 TACACCACTGATGATCTTGTAAACAAACATACAGACCTAATTTCTACTAGAAAAAC 28479
Qy 1585 ----- 1584
Db 28478 CTGTTGTGACTTACCAACTCTTATTTCTTGTGACTTCTGTGAATCTGCTGATCTAAC 28419
Qy 1585 -----TTCCTTGAGAACCTGGAAGTGGTATCCTCCACACATC 1625
Db 28418 AAGATGTTGCTTATTTATTTAGTTCCTTGAGAACTGGAAGTGGTATCCTCCACACATC 28359
Qy 1626 AATGTTAGTGTGATGAAGAAAGTGGCAAAATGCTGATCCAGAGAAAGCAAGAAACTG 1685
Db 28358 AATGTTAGTGTGATGAAGAAAGTGGCAAAATGCTGATCCAGAGAAAGCAAGAAACTG 28299
Qy 1686 GCAGAAATTTGTCAGATGCAAAATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1745
Db 28298 GCAGAAATTTGTCAGATGCAAAATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 28239
Qy 1746 AGCTAAAGACCACTGGATGTTTATCTACGAGTCAATTAAGTCTTGTGCTTACTTC 1805
Db 28238 AGCTAAAGACCACTGGATGTTTATCTACGAGTCAATTAAGTCTTGTGCTTACTTC 28179
Qy 1806 TGAAA----- 1810
Db 28178 TGAAAAGTATATTTCCCTTGTGAGATGATTTGCTTAACATATTTTACATGAAGGGG 28119
Qy 1811 -----AGTGGGCTTACATGCTAGCAAC 1835
Db 28118 CAGTTTGGTCAATTTGACTTTTGTATCTCATGTTTATAGTGGGCTTACATGCTAGCAAC 28059
Qy 1836 AATCAAGAGCAATTTATTAAGAAGCTTTAGAGCAAGAGAGTATGTTGAGATGAG 1895
Db 28058 AATCAAGAGCAATTTATTAAGAAGCTTTAGAGCAAGAGAGTATGTTGAGATGAG 27999
Qy 1896 AATCTTATGCTGATGAGTGGTGAAGCGGCTAGCGCTTC----- 1934
Db 27998 AATCTTATGCTGATGAGTGGTGAAGCGGCTAGCGCTTCGGGTAGATATTTTGTCTGTT 27939
Qy 1935 ----- 1934
Db 27938 TTGATAGTTGCTTCGATCTCAGAGCAATTTTGTGATGATTTACTTACTGCTGTTGAT 27879
Qy 1935 ----- 1934

```



Db 301 GAGCGTCACAGTAAGAGAGACAGCGATTGTGTAGATGACGCGGCAACGTGT 360  
 Oy 421 CGATCCGCTGCTGATGACGAGGAATAGCTTTGATCGGCTTTGCATCCAAATCGGAGT 480  
 Db 361 CGATCCGCTGCTGATGACGAGGAATAGCTTTGATCGGCTTTGCATCCAAATCGGAGT 420  
 Oy 481 TTGAGGATCTAAATCGGAATTTGCAATTAATCTCATCTCCATCTCTCTGAGAGTCCG 540  
 Db 421 TTGAGGATCCAAATCGGGAATTTGCAATTAATCTCATCTCTCTGAGAGTCCG 480  
 Oy 541 AATCGATCTACACACACTACTCGTACCGCGGCTCATTTACTGCCCGGATTTCAAAATTA 600  
 Db 481 AATCGATCTACACACACTACTCGTACCGCGGCTCATTTACTGCCCGGATTTCAAAATTA 540  
 Oy 601 TCCGATCATTTCCGGCGATATCCAAATCGACACTGA 636  
 Db 541 TCCGATCATTTCCGGCGATATCCAAATCGACACTGA 576

## RESULT 7

: us-10-219-999-9087  
 : Sequence 9087, Application US/10219999  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Edgerton, Michael D  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Kovalic, David K.  
 : APPLICANT: Liu, Jindong  
 : APPLICANT: Stein, Joshua  
 : TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
 : FILE REFERENCE: 38-10(52726)C  
 : CURRENT APPLICATION NUMBER: US/10/219, 999  
 : PRIOR FILING DATE: 2002-08-15  
 : PRIOR APPLICATION NUMBER: US 60/324,109  
 : PRIOR FILING DATE: 2001-09-21  
 : PRIOR APPLICATION NUMBER: US 60/312,544  
 : PRIOR FILING DATE: 2001-08-15  
 : NUMBER OF SEQ ID NOS: 63520  
 : SEQ ID NO 9087  
 : LENGTH: 1956  
 : TYPE: DNA  
 : ORGANISM: Zea mays  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (140)..(1675)  
 : OTHER INFORMATION:  
 : US-10-219-999-9087

Query Match 21.6%; Score 518.4; DB 42; Length 1956;  
 Best Local Similarity 61.0%; Pred. No. 1.4e-123;  
 Matches 941; Conservative 0; Mismatches 571; Indels 30; Gaps 5;

Oy 685 ATGGCTGTGTTGCTTCTGCTGCTCGGGAAGTTTGATGACGAGGCTACTCTTACTC 744  
 Db 140 ATGGAGGAGGTGGCTGCGCGCGGGAAGTGTCTATCGCGGAGATACCTCTGCTG 199  
 Oy 745 GAGAAGCAAAATGACGGCTTGTGTTGATGACAAATGCAACGTTTACGATTTGAAG 804  
 Db 200 GAGCGCCCAACGCGCGGCTAGTCTCTGACACACCGCCGCTTCTAGCCCTCGTCCG 259  
 Oy 805 CCAATCAACGAAGAAGTCAAGCCTGAAGTTGGCATGGAATGACAGATGCAAAATTA 864  
 Db 260 CCCCCTCCCGACAGCCTCCCGCCGACGCTGGGCTGGGCTGAGCGATGCAAAAGTG 319  
 Oy 865 ACATCACACAGCTCTGAGAGAAAGCANTGATAAGTCTCAATCATTTGACTCTT 924  
 Db 320 ACTTCCTCAGCTCTCTACGGGAGGCCACATACAGCTGTCCCTTAACGAAGTCAGCGTG 379  
 Oy 925 CAGTGTGTCTGCAAGATGATTCAGAAACCCCTTTAGAGCATGACATACAGTATGCT 984  
 Db 380 CAATTGACATCTCGAGGAGGAATCTCAAAACCTTTTGTGAGCAGGCCATACAAATTTTCG 439  
 Oy 985 ATAGCTGTGCTCATTTGGCAA-----CCGAGAAAGAACAAAGATCATTTGCACAAACTC 1038

Db 440 GTAGACGCTGCTAAAGCCGACCATTTATGATAGAGAGAGAAAGATGATAGATTAAGTTG 499  
 Oy 1039 TTATTGCAAGGCTCTGATATATACATATTTAGCTCCCAATGATCTTTACTCATATGGGAA 1098  
 Db 500 CTGCTCAAGGCTCTGATATATACATATTTAGCTCCCAATGATCTTTACTCATATGGGAA 559  
 Oy 1099 CAGATAGATGCGGTGGGCTTCATTGACACCGAGATGCGTGGTACCCTTGACCGGTTT 1158  
 Db 560 CAGATTGAAGACGCGGCTCTCCCTCTTACTCCAGAGGCACTCTTCAATGGCTCATTC 619  
 Oy 1159 GCATCAATCACATTCATTAATGCTGGAGTCAAAATGCTCT-----AATTCGAG 1206  
 Db 620 TCTTCATATCATTAACTCAAGAGTGTGATGAAACAAATGACTGAGAGAAATGTA 679  
 Oy 1207 CCTGAAGTAGCAAAACGCTGTTAGTCTTCTGACAGCAATACAAACGCTGTTGCA 1266  
 Db 680 CCTGAAGTAGCCAAACGCTGTTAGTCTTCTGACAGCAATACAAACGCTGTTGCA 739  
 Oy 1267 GCTGTGTACATTAATCTGAGAGTGGTGGCTGACATCTCATCTGATCCATGTA 1320  
 Db 740 GCTCTTCTTCACTATCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 799  
 Oy 1321 AAGTTGGCTGTTCTGATCTAGATGTTATCATATGATAGACAAACGCTGCTGCTGCT 1380  
 Db 800 AATACACTAGACGAGAACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 859  
 Oy 1381 GCACAAAGGAGCTGGGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1440  
 Db 860 GCACAGGGGAAATTTGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 919  
 Oy 1441 TATGTTGCTGCTCTCCAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 Db 920 TATGTAAGTTTCTCCAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976  
 Oy 1501 TTAATGAAATTTATGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTA 1560  
 Db 977 CCAGATGTATGATCAAT---ATTGTTACCAAAAGTGGAGATCAATGAAATTAAGATTTC 1033  
 Oy 1561 TCTTTACCCACCATGATGATGATTTCTTCTGGAAGAACTGGAGAGTGGATCTCCACA 1620  
 Db 1034 TCAATTAACCTCTCTGATGACCTCTCTCTGGAAGAACTGGAGAGTGGATCTCCACA 1093  
 Oy 1621 CCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
 Db 1094 CCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153  
 Oy 1681 AACTGGCAGATTTGTCAGATCAAAATTTAGAACTGGAACTTAACGATCTGAGC 1740  
 Db 1154 ACATGGAGTAAACTGCGCTTCTCCCAATTCAGCGCTGGAGAACCACTGGCAATCTTAA 1213  
 Oy 1741 AATTAAGTAAAGACCATGAGGATGTTATCTACAGAGTATTAAGTCTGATGCTGCTT 1800  
 Db 1214 GGAATTTCTGAAATACCGAGAGGATATGATGATGATGATGATGATGATGATGATGAT 1273  
 Oy 1801 ACTTCGAAAGTGGGTGTTTACATGCTACTGACCAATCAACGAAGCCATTAATTAAGAA 1860  
 Db 1274 ACTTATGAGAAAGGAGCAAGGAGGCTACTAACAACATCAAGAAATTAATTAATTA 1333  
 Oy 1861 CTCTTAGAGGCAAGAGACTATGTTGAGATCAGAAATCTTATGCTGATAGTGGTGAG 1920  
 Db 1334 TTGTTGGCTGCAAGGAGTCTTGGCTTGAAGTAAGGCTTCAATGCGAAGATGGGCATA 1393  
 Oy 1921 GGGGCTAGGCTCCGATGAGCTGAATCTCAACCTTTGGATCTCTCAATGAGT 1980  
 Db 1394 GCAGCGGCTTCCGATGAGCTGAGATGATACAAACGCTTCTGACCCCAATGAAAT 1453  
 Oy 1981 GCTGAAGAGTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040  
 Db 1454 ATGAGAGGCTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1513  
 Oy 2041 ATCACTTATGAGGATTCGGGACCAAACTGACCCAGGATGAGTGGCAATGTTTTC 2100











OY	1681	AAGGCGAAGATTTCACAGTATGCAAATTTAGAATCGGAACAATAACTTAAGACTGTGAC	1740
Db	1123	ACATGGAGTAACTGGCCTTTGCCAATTTAGCGTGGAACAACCAATCCGAATCTTAAA	11822
OY	1741	AAATTAACCTAAAGCACACTGGCATGTTTTATACGAGTCCATTAASTCTTGAATGTCTT	18009
Db	1183	GGAATTTCTGAANAATCACCGAAGG-----	1207
OY	1801	ACTTCTGAAAAAGTGGGTGTTACATGCTACTGAACCATAATCAAGAACCATTAATAAGAA	18609
Db	1208	-CATATGAGTCGTGGACAGAGGGTGCTACTAACCAACATCAGAAATTAATTATTAATCA	12666
OY	1861	CTCTTAGAGCAAAGAGAAGTATGTTGAGAGATCAGAAATTCATTATGCGTCAGATGGGTAG	19209
Db	1267	TGTGTGCTCCAAAGGAGATGCTTTGCTCTTGAGTAAGGCTTCACATCCGAGAAATGGGCATA	13266
OY	1921	GGGGCTAGGCTTCGCGATAGAGCCTGAATCTCAAACTCAACTTTTGAGATTTCATCATGAT	19809
Db	1327	GCACCCGGTTCGCGANTGAGCCAGATTCACAAACACGGCTACCTTGACGCCACATATGAT	13866
OY	1981	GCTTAAGAGATTTACATGGTGTGTTCCGTGAGAGCGTGGAGATTGATGCCATATTTGCA	20409
Db	1387	ATGGAAGGATTTACTTCTGCTGGAGTTCCGGAGCTGGTGGCTTGATGATGCATATTTCCA	14466
OY	2041	ATCATTTTAGGGATTTCCGGCACCAAACTGACCCAGGCAATGAGTTCCACAAATGTTTTG	21009
Db	1447	GTCAGTTTGGGGGATGCAAGTATGCTGTAGCAAAATGCTTGGAGCTCTGCTGCGCTTGTG	15066
OY	2101	GCTTTGTTGGTAGAGAGATCCACATGSCGTTTGGCTTAGAAGAAAGTGATGCCAGAAC	21609
Db	1507	CCAACTTCGTTGAGAGAAAGTCCCGTGGTGTCTCATTTGGAAGATGCTGACCCAAAGAA	15666
OY	2161	ACATGTTTACTTCAGGCTTTCATCAATTCACCTTGAATA 2202	
Db	1567	AGGAGGTATCGCAGCTGTATGTTCCATTCAGATTAAATGA 1608	

	RESULT	12	
	US-09-638-111-871/C		
	; Sequence	871 Application US//0968111	
	; GENERAL INFORMATION:		
	; APPLICANT:	G <sup>a</sup> Lach, J <sup>r</sup> rn	
	; TITLE OF INVENTION:	Expressed sequences of Arabidopsis	
	; TITLE OF INVENTION:	thaliana	
	; FILE REFERENCE:	200805	
	; CURRENT APPLICATION NUMBER:	US//09/638,111	
	; CURRENT FILING DATE:	2000-08-11	
	; PRIOR APPLICATION NUMBER:	US 60/148,783	
	; PRIOR FILING DATE:	1999-08-13	
	; NUMBER OF SEQ ID NOS:	900	
	; SOFTWARE:	FastSeq for Windows Version 3.0	
	; SEQ ID NO	871	
	; LENGTH:	443	
	; TYPE:	DNA	
	; ORGANISM:	Arabidopsis thaliana	
	US-09-638-111-871		
OY	1896 AATTCCTTATGGCTCAGATGCGGTGAGGCGGCTTAGACGCTCCGATAGAGCCTGAATCTCAAAC	18.5%; Score 443; DB 24; Length 443;	
	Best Local Similarity	100.0%; Pred. No. 3.4e-104;	
	Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Db	443 AATCTTAATGGCTAGAGAGGCTGAGGCGGCTTAGACGCTCCGATAGAGCCTGAATCTCAAAC		384
OY	1956 TCAACTTTTGAGATTCTACATGATGAGTGCCTGAAGAAGATTCTACTCTGCTGGTGTCTCTGGAGC		2015
Db	383 TCACTTTTGAGATTCTACAATGATGAGTGCCTGAAGAAGATTCTACTCTGCTGGTGTCTCTGGAGC		324
OY	2016 TGCTGAGATTGATGCCAATTTTGCATCTACTTACGGGATTCGGCACCAAACTACCCA		2075
Db	323 TGCTGAGATTGATGCCAATTTTGCATCTACTTACGGGATTCGGCACCAAACTACCCA		264

Qy	2076	GGATGAGATGTCACAAATGTTTGGCCCTGTTGGTGAGAGAAGAAATCCAAATGGCGTTG	2135
Db	263	GGCATGAGATTCGCACATGTTTTGGCTGTGGTGAGAGAAGAAATCCAAATGGCGTTTG	204
Qy	2136	CCTGAAAGTGATGATCCAGAACCAATGATATTACTTCAGCGTTTATCATATCACT	2195
Db	203	CCTGAAAGTGATGATCCAGAACCAATGATATTACTTCAGCGTTTATCATATCACT	144
Qy	2196	TGACATAACAACATGTTTCAGTCCATTTATAGTCCGCACCAATGCGTGGAGT	2255
Db	143	TGACATAACAACATGTTTACGTCCATTTATAGTCCGCACCAATGCGTGGAGT	84
Qy	2256	ATACTGTTTGCATATAGACTTGGGTCTAAATTTCTGGTGAAGCAATTTTATATCCA	2315
Db	83	ATACTGTTTGCATATAGACTTGGGTCTAAATTTCTGGTGAAGCAATTTTATATCCA	24
Qy	2316	TTGTAAAGCTCTTAACTCTTGA	2338
Db	23	TTGTAAAGCTCTTAACTCTTGA	1

```

RESULT 13
US-09-770-444-851/c
; Sequence 851, Application US/09770444
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameeka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 851
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-851

```

	Query Match	Similarity	18.5%	Score 443	DB 30	Length 443
	Best Local	Similarity	100.0%	Pred. No. 3.4e-104		
	Matches 443	Conservative	0	Mismatches	0	Indels 0; Gaps 0;
QY	1896	AATTCCTAATGCGTCGATGCGGCGGCTAAGCCTTCGATAGAGCCTGAATCTCAAC	1955			
Db	443	AATTCCTAATGCGTCGATGCGGCGGCTTCGATAGAGCCTGAATCTCAAC	384			
QY	1956	TCAACTTTTGAGTTCTACAAATGAGTGGCGAAGAGATTCCTACTTGCGTCCCTGGAGC	2015			
Db	383	TCAACTTTTGAGTTCTACAAATGAGTGGCGAAGAGATTCCTACTTGCGTCCCTGGAGC	324			
QY	2016	TGCTGATTTGATGCGCAATATTTGCATCACTTTAGGGGATTCGGACACCAACTGACCA	2075			

Db 323 TGGTGGATTGATGCCAATTTGCAATGACTTTAGGGGATTCGGGACCAAACTGACCCA 264  
Qy 2076 GGCAATGAGTTGCGACAAATGTTTGGCCTGTGGTGAGAGAAGATCCACAATGGCGTTTG 2135  
Db 263 GGCAATGAGTTGCGACAAATGTTTGGCCTGTGGTGAGAGAAGATCCACAATGGCGTTTG 204  
Qy 2136 CCTGAAGAAGTGTGATCCAGACACATGTAATTAATCTTCAGCGCTTTTATCAATTCACCT 2195  
Db 203 CCTGAAGAAGTGTGATCCAGACACATGTAATTAATCTTCAGCGCTTTTATCAATTCACCT 144  
Qy 2196 TGAGTAACAACATTTGTTGAGTGTGCAATTAATAGTGTGCGTCCACCAAGTTCGGTTGAGT 2255  
Db 143 TGAGTAACAACATTTGTTGAGTGTGCAATTAATAGTGTGCGTCCACCAAGTTCGGTTGAGT 84  
Qy 2256 ATACTGTTTGCATATAGACTGGGTGCTAAATTTCTTGGTGAAGCATTTTATATACCA 2315  
Db 83 ATACTGTTTGCATATAGACTGGGTGCTAAATTTCTTGGTGAAGCATTTTATATACCA 24  
Qy 2316 TTGTAGGCTCTTAACTCTTGA 2338  
Db 23 TTGTAGGCTCTTAACTCTTGA 1

## RESULT 14

US-09-654-617-192867  
; Sequence 192867, Application US/09654617  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jindong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 38-21(15097)D  
; CURRENT APPLICATION NUMBER: US/09/654,617  
; CURRENT FILING DATE: 2000-09-05  
; NUMBER OF SEQ ID NOS: 463173  
; SEQ ID NO 192867  
; LENGTH: 728  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; OTHER INFORMATION: unsure at all n locations  
US-09-654-617-192867

Query Match 17.88; Score 425.4; DB 25; Length 728;  
Best Local Similarity 76.18; Pred. No. 1.6e-99;  
Matches 554; Conservative 0; Mismatches 162; Indels 12; Gaps 2;

Qy 784 CGGTTTACGGGATTTGAGACCAATCAACGAGAGTCAAGCCTGGAAGTTGGGCATGG 843  
Db 1 CGTTTTCAGCTATTTGTAAGCAATTCATGAGCTATCAAGCCTGGAAGCCTGGGCATGG 60  
Qy 844 AAATGACAGATGTCAAATTAACATCACACAGCTCTCGAGAGAAAGCATGTATAACTG 903  
Db 61 TCTTGACCGATGTCACCACTCACTCTCAGCTTCCAGAGAAAGCATGTATAACTG 120  
Qy 904 TCACGATCATTTGAGCTCAGTCTGCTGCTGCAAGTGAATCAAGAAACCCCTTTGTA 963  
Db 121 TCTGGAACATTTAATCACTCAAGTGTATCTTCAAGTGAATCAAGAAACCCCTTTGTA 180  
Qy 964 GAGCATGCGATACAGTATGATAGCTGCTGCTCATTTGGCA---ACGAGAGAGACAAA 1020  
Db 181 GAAATGCTATTAATATACATAGACAGCTGACATGCAACATTTGCAAAAGATAGAAA 240  
Qy 1021 GAATCATGACAAACTCTTATTCGAAGCTTGTATATACAAATTAAGCTTCACATGAC 1080  
Db 241 GAGCATGATAGTAATCACTCTTCAAGAGCTTGTATATAGATCTTGTGCAATGAC 300  
Qy 1081 TTTTACATATTCGGAACCAATAGAAATCGGCTGGCTTCATGACACCGAATCGCTG 1140  
Db 301 TTTTACATATTCGGAACCAATAGAAATCGGCTGGCTTCATGACACCGAATCGCTG 360  
Qy 1141 GGTACCCCTGGACCGTTGATCATATTCATGATTCATGATGCTGGGAGTCAAAATGGCTAT 1200  
Db 361 GGTACCTTACACCGCTTATCATATTAATTCATATTTCTGAGGAATCAAAATGAGCAAT 420

Qy 1201 TCCAAAGCTGAAGTAGCAAAAAGTGGCTTGGATCTCTGACAGCATGACAAAGCTGTG 1260  
Db 421 TGCAAAAGCTGAAGTAGCAAAAAGTGGATCTCTGACAGCATGACAAAGCTGTGTA 480  
Qy 1261 GTTGAGCTGCTTACATTAATCTGGATGGTGGACCTATCGAATCCATG---AAAG----- 1315  
Db 481 GTTGAGCTGCTTACATTAATCTGGATGGTGGACCTATCGAATCCATG---AAAG----- 1315  
Qy 1316 ---AAGAAAGTTGGCTGTTGATCTGATAGATGTTATCAATATGATAGCAAAAGCTCT 1371  
Db 541 CACCAAGAAATTAAGAAATTCACAGATCTGATATGATAGATAGATAGCAAAAGCTGCC 600  
Qy 1372 CATGCTCTGCAACAAGGAGAGTGGAGTGGCTTGAATGTCAGCTGCTGTATGGA 1431  
Db 601 CACTGATATGCGCAAGGATGAAGTGGAGTGGCTTGAATGTCAGCTGCTGTATGGA 660  
Qy 1432 AGTCAGCTTATGTTGCTGCTGCTGCAAGCTGCTGATATTTGCTCAGCTGCAAGTACA 1491  
Db 661 AGTCAGCTTATGTTGCTGCTGCTGCAAGCTGCTGATATTTGCTCAGCTGCAAGTACA 720  
Qy 1492 GGTCTGCC 1499  
Db 721 GGGATGCC 728

## RESULT 15

US-09-684-016-192867  
; Sequence 192867, Application US/09684016  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jindong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 38-21(15097)D  
; CURRENT APPLICATION NUMBER: US/09/684,016  
; CURRENT FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: US 09/654,617  
; PRIOR FILING DATE: 2000-09-05  
; NUMBER OF SEQ ID NOS: 463173  
; SEQ ID NO 192867  
; LENGTH: 728  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(728)  
; OTHER INFORMATION: unsure at all n locations  
US-09-684-016-192867

Query Match 17.88; Score 425.4; DB 27; Length 728;  
Best Local Similarity 76.18; Pred. No. 1.6e-99;  
Matches 554; Conservative 0; Mismatches 162; Indels 12; Gaps 2;

Qy 784 CGGTTTACGGGATTTGAGACCAATCAACGAGAGTCAAGCCTGGAAGTTGGGCATGG 843  
Db 1 CGTTTTCAGCTATTTGTAAGCAATTCATGAGCTATCAAGCCTGGAAGCCTGGGCATGG 60  
Qy 844 AAATGACAGATGTCAAATTAACATCACACAGCTCTCGAGAGAAAGCATGTATAACTG 903  
Db 61 TCTTGACCGATGTCACCACTCACTCTCAGCTTCCAGAGAAAGCATGTATAACTG 120  
Qy 904 TCACGATCATTTGAGCTCAGTCTGCTGCTGCAAGTGAATTAAGAAACCCCTTTGTA 963  
Db 121 TCTGGAACATTTAATCACTCAAGTGTATCTTCAAGTGAATCAAGAAACCCCTTTGTA 180  
Qy 964 GAGCATGCGATACAGTATGATAGCTGCTGCTCATTTGGCA---ACGAGAGAGACAAA 1020  
Db 181 GAAATGCTATTAATATACATAGACAGCTGACATGCAACATTTGCAAAAGATAGAAA 240  
Qy 1021 GAATCATGACAAACTCTTATTCGAAGCTTGTATATACAAATTAAGCTTCACATGAC 1080  
Db 241 GAGCATGATAGTAATCACTCTTCAAGAGCTTGTATATAGATCTTGTGCAATGAC 300  
Qy 1081 TTTTACATATTCGGAACCAATAGAAATCGGCTGGCTTCATGACACCGAATCGCTG 1140

```
Db 301 TTTACTCATACAGATCAGATAGAACACTTGGTCTTCGGTTGACACCTGAACCATTTG 360
QY 1141 GGTACCCCTTGACACCGTTTGACATCAATCAATTCATTCGCGAGTCAAAATGCTCAAT 1200
Db 361 GCTACTCTACACCGTTTGACATCAATTCATTCATTCGAGATCAAAATGAGCAAAAT 420
QY 1201 TCCAGACCTGAAGTAGCAAAACAGCGCTTAGTCTCTGAGCAATGACAAACAGCTGAG 1260
Db 421 TGCAAACCTGAAGTTGCAAAACAGCGATTCATCTGACGCAATGACAAACAGCTGTA 480
QY 1261 GTTCAGCTCTGTACATTAATCTGAGAGTGGTGAATCTGATCATGATTAAG----- 1315
Db 481 GTTGTGCTTACTTCATTAATCTGAGTGTGTTAACTTCACCTCTTCGACATCA 540
QY 1316 ----AAGGAAGTTGGCTGTCTGATCTAGATGTTATCAATGATGACAAAGCTCT 1371
Db 541 CACCAAGAAATAGAAATTCACAGATCTGATATGATGATGATGATGATGATGATGATG 600
QY 1372 CATGCTCTTGACAAAGGAGGTGAGGTGATGATGATGATGATGATGATGATGATGATG 1431
Db 601 CACTGTATTCGCCAAGGATTAAGTTGGAGTGGCTTTGATGATGATGATGATGATGATG 660
QY 1432 AGTCAGCGTATGTTGCTCTCTCCAGAGTCTTGCATTTGCTCAGAGTTGCACTAACA 1491
Db 661 AGTCAGCGTATGTTGCTCTCTCCAGAGTCTTGCATTTGCTCAGAGTTGCACTAACA 720
QY 1492 GGTCTGCC 1499
Db 721 GGGATGCC 728
```

Search completed: May 3, 2003, 21:11:50  
Job time : 3833 secs

THIS PAGE BLANK (USPTO)

[illegible]

|||||  
Db 241 GTTCTGCTTTCTCCGAACTCCAGCCAGTTGGTTTATTTTTCACAGATTTGTC 300  
OY 301 TTCTCTTTTCATGGGCGACGCGACGTCGATTTCTGTCAAAAGCTTAACATCCGACTC 360  
Db 301 TTCTCTTTTCATGGGCGACGCGACGTCGATTTCTGTCAAAAGCTTAACATCCGACTC 360  
OY 361 GAGCGTCACAGTAAGAGAAAGACAGCAATGTGTGTAGATGACGGCGACAGCTGTGT 420  
Db 361 GAGCGTCACAGTAAGAGAAAGACAGCAATGTGTGTAGATGACGGCGACAGCTGTGT 420  
OY 421 CGATCCGTCCTGATCGACGAGAAATAGCTTTCGATCCGCTTTCATCCAAATGCGAGCT 480  
Db 421 CGATCCGTCCTGATCGACGAGAAATAGCTTTCGATCCGCTTTCATCCAAATGCGAGCT 480  
OY 481 TTGAGATCTTAATCGAAATTTGCAATTAATCTCATCTCCAAATCTCTCTGAGAGTCG 540  
Db 481 TTGAGATCTTAATCGAAATTTGCAATTAATCTCATCTCCAAATCTCTCTGAGAGTCG 540  
OY 541 AATCCGATCTACACCACTACTGCTACCGCGGTCAATTTACTGCGCGCATTTCAATTA 600  
Db 541 AATCCGATCTACACCACTACTGCTACCGCGGTCAATTTACTGCGCGCATTTCAATTA 600  
OY 601 TCCGATATTTCCGCGCATATCCCAATGCGACACTGAGGTGAATCTGGGGTTTGAATACG 660  
Db 601 TCCGATATTTCCGCGCATATCCCAATGCGACACTGAGGTGAATCTGGGGTTTGAATACG 660  
OY 661 GATATCTTTTGTCACTCTTTGAAATGAGCTTTGCTTTCGCTTCCGCGGAAAGTTTG 720  
Db 661 GATATCTTTTGTCACTCTTTGAAATGAGCTTTGCTTTCGCTTCCGCGGAAAGTTTG 720  
OY 721 ATGACGTGAGGCTACTCTTGTACTCGAGAGCCAAATGCAAGGCTTGTGTGATCAAT 780  
Db 721 ATGACGTGAGGCTACTCTTGTACTCGAGAGCCAAATGCAAGGCTTGTGTGATCAAT 780  
OY 781 GCAGGCTTTTACGCGATTTGTAAGCCAAATCAAGCAAGAAAGTCAAGCTTGAAGTTGGCA 840  
Db 781 GCAGGCTTTTACGCGATTTGTAAGCCAAATCAAGCAAGAAAGTCAAGCTTGAAGTTGGCA 840  
OY 841 TGGAAATGACAGATGTCAAAATTAACATCAACACAGCTCTCGAGAGAAAGCATGTATA 900  
Db 841 TGGAAATGACAGATGTCAAAATTAACATCAACACAGCTCTCGAGAGAAAGCATGTATA 900  
OY 901 CTGTCACTGAATCAATTTGACTCTTCACTGTGTGTGTGCAAGTGAATCAGAAACCCCTT 960  
Db 901 CTGTCACTGAATCAATTTGACTCTTCACTGTGTGTGTGCAAGTGAATCAGAAACCCCTT 960  
OY 961 GTAGAGCATGCGATACAGATGATGATAGCTGTGCTCATTTGGCAACCGAAGAGACAAA 1020  
Db 961 GTAGAGCATGCGATACAGATGATGATAGCTGTGCTCATTTGGCAACCGAAGAGACAAA 1020  
OY 1021 GAATCATTTGCAAAACTCTTATTTGCAAGGCTTGTATATTAACAAATTAAGCTCCCAATGAC 1080  
Db 1021 GAATCATTTGCAAAACTCTTATTTGCAAGGCTTGTATATTAACAAATTAAGCTCCCAATGAC 1080  
OY 1081 TTTTATCTATATGCGAAGCAGATAGAAATGCGCTGGCTTCCATTGACACCAATTCGCTG 1140  
Db 1081 TTTTATCTATATGCGAAGCAGATAGAAATGCGCTGGCTTCCATTGACACCAATTCGCTG 1140  
OY 1141 GGTACCCCTTGACCGCTTGTGCAATCAATCAATGCTGGGAGTCAAAATGGTGTAT 1200  
Db 1141 GGTACCCCTTGACCGCTTGTGCAATCAATCAATGCTGGGAGTCAAAATGGTGTAT 1200  
OY 1201 TCCAAAGCTGAAGTAGCAAAAACGCGTTAGTTCTTCTGACAGCAATGACAAAGCTGTG 1260  
Db 1201 TCCAAAGCTGAAGTAGCAAAAACGCGTTAGTTCTTCTGACAGCAATGACAAAGCTGTG 1260  
OY 1261 GTTCACAGCTCTGTACATTAATCTTGAGAGTGTGACCTATCTGATTAAGAAAGGA 1320  
Db 1261 GTTCACAGCTCTGTACATTAATCTTGAGAGTGTGACCTATCTGATTAAGAAAGGA 1320  
OY 1321 AAGTTGGCTGTGATCTGATGATGATATCCATATGATAGCAAAAGCTCATTTGCTT 1380  
|||||

Db 1321 AAGTTGGCTGTGATCTGATGATGATATCCATATGATAGCAAAAGCTCATTTGCTT 1380  
OY 1381 GCACAGAGGAAGTCGGAAGTGGCTTGTATGATGTACGTGTGCTGTATGGAAGTACAGCT 1440  
Db 1381 GCACAGAGGAAGTCGGAAGTGGCTTGTATGATGTACGTGTGCTGTATGGAAGTACAGCT 1440  
OY 1441 TATGTGCTTCTCTCCAGAAATGCTTGTATTTGCTGAGTTGAGTATACAGTGTGCCA 1500  
Db 1441 TATGTGCTTCTCTCCAGAAATGCTTGTATTTGCTGAGTTGAGTATACAGTGTGCCA 1500  
OY 1501 TTAATAGATTAATTTGTATCAATTTTGAAGGAAATGAGCAATTAAGAGACATGATCTC 1560  
Db 1501 TTAATAGATTAATTTGTATCAATTTTGAAGGAAATGAGCAATTAAGAGACATGATCTC 1560  
OY 1561 TCTTTACACACATGATGAATCTTTTCCCTTGAGAACCTGGAAGTGTGATCTCCACA 1620  
Db 1561 TCTTTACACACATGATGAATCTTTTCCCTTGAGAACCTGGAAGTGTGATCTCCACA 1620  
OY 1621 CCATCAATGTTAGTGCATGAAGAGTGGCAAAATGCTGTATCCAGAGAGGCGACAGAA 1680  
Db 1621 CCATCAATGTTAGTGCATGAAGAGTGGCAAAATGCTGTATCCAGAGAGGCGACAGAA 1680  
OY 1681 AACTGCGAGATTTGTGAGATGCAAAATTTAGACTGGAATTAAGCTTAAGATCTGAGC 1740  
Db 1681 AACTGCGAGATTTGTGAGATGCAAAATTTAGACTGGAATTAAGCTTAAGATCTGAGC 1740  
OY 1741 AAATTTAGCTAAAGACCACTGGGATGTTTATCTACAGATCAATTAAGCTTGTAGTGTCTT 1800  
Db 1741 AAATTTAGCTAAAGACCACTGGGATGTTTATCTACAGATCAATTAAGCTTGTAGTGTCTT 1800  
OY 1801 ACTTCTGAAAAGTGGGTGTATCATGCTATGTAACCAATCAACGAAGCATTTATTAAGAA 1860  
Db 1801 ACTTCTGAAAAGTGGGTGTATCATGCTATGTAACCAATCAACGAAGCATTTATTAAGAA 1860  
OY 1861 CTCTTAGAGGCAAGAGAAAGTATGTTGAGATGCAAAATTTCTATGCGCTAGATGAGT 1920  
Db 1861 CTCTTAGAGGCAAGAGAAAGTATGTTGAGATGCAAAATTTCTATGCGCTAGATGAGT 1920  
OY 1921 GCGGCTAGCGCTTCCGATAGAGCTGAATCTCAAACTCAACTTTTGGATTTCAATAGT 1980  
Db 1921 GCGGCTAGCGCTTCCGATAGAGCTGAATCTCAAACTCAACTTTTGGATTTCAATAGT 1980  
OY 1981 GCTAAGAGATTTCTACTTGTGCTGTGCTTCTGAGCTGTGATTTGATGCCATTTTGA 2040  
Db 1981 GCTAAGAGATTTCTACTTGTGCTGTGCTTCTGAGCTGTGATTTGATGCCATTTTGA 2040  
OY 2041 ATCACTTTAGGGGATTCGCGACCAAACTGACCCAGGATGAGTTGCGCAAAATGTTTG 2100  
Db 2041 ATCACTTTAGGGGATTCGCGACCAAACTGACCCAGGATGAGTTGCGCAAAATGTTTG 2100  
OY 2101 GCGTTGTTGTGAGAGAAATGATCATGCGCTTTCCTAGAAAGTGTGATCCAGAAC 2160  
Db 2101 GCGTTGTTGTGAGAGAAATGATCATGCGCTTTCCTAGAAAGTGTGATCCAGAAC 2160  
OY 2161 ACATGATATTTACTAGGCGCTTTCATCAATTCACCTTGAATTAAGCAATGTTTCACTGT 2220  
Db 2161 ACATGATATTTACTAGGCGCTTTCATCAATTCACCTTGAATTAAGCAATGTTTCACTGT 2220  
OY 2221 CCAATTTATAGTGTGCTGACCAAACTGCTGAGTATAGCTTTTTCGATATAGACTGTGG 2280  
Db 2221 CCAATTTATAGTGTGCTGACCAAACTGCTGAGTATAGCTTTTTCGATATAGACTGTGG 2280  
OY 2281 TGCTAAATTTCTTGTGTAGCAATTTTATACCATTTGAAGTCTTTAACTCTTGAAA 2340  
Db 2281 TGCTAAATTTCTTGTGTAGCAATTTTATACCATTTGAAGTCTTTAACTCTTGAAA 2340  
OY 2341 ACTTGGCGGAAAAATTAAGTGTGATTTCAATCTTTCAAAAAAATTTTTTTTTTTTTT 2396  
Db 2341 ACTTGGCGGAAAAATTAAGTGTGATTTCAATCTTTCAAAAAAATTTTTTTTTTTTTT 2396  
|||||

RESULT 2  
US-09-770-444-851/c



Sequence 851, Application US/09770444  
Patent No. US20020023280A1  
GENERAL INFORMATION:  
APPLICANT: Gorlach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Mathew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Moessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kicker, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2027 (PARA-016PRV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
PRIORITY FILING DATE: 2001-01-26  
PRIORITY APPLICATION NUMBER: 60/178,502  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 851  
LENGTH: 443  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-851

Query Match 18.5%; Score 443; DB 10; Length 443;  
Best Local Similarity 100.0%; Pred. No. 2,8e-121;  
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1896 AATCTTATGCGTCAGATGGGTGAGCGCGCTAGCGCTTCGATAGAGCCGTAATCTCAAC 1955  
DB 443 AATCTTATGCGTCAGATGGGTGAGCGCGCTAGCGCTTCGATAGAGCCGTAATCTCAAC 384  
QY 1956 TCACTTTTGGATTTCAACAATGAGTGTGAGAGGAGTTCTACTTGTGTGCTTCTGAGAGC 2015  
DB 383 TCACTTTTGGATTTCAACAATGAGTGTGAGAGGAGTTCTACTTGTGTGCTTCTGAGAGC 324  
QY 2016 TGGTGATTTGATGCGCATATTTGCAATCACTTAGGGGATTCGGGACCAACCTGACCA 2075  
DB 323 TGGTGATTTGATGCGCATATTTGCAATCACTTAGGGGATTCGGGACCAACCTGACCA 264  
QY 2076 GGCATGAGGTTGCGACAAATGTTTGGCTGTGTGAGAGAGATTCACATGCGGTTTG 2135  
DB 263 GGCATGAGGTTGCGACAAATGTTTGGCTGTGTGAGAGAGATTCACATGCGGTTTG 204  
QY 2136 CCTAAGAGTGTGATTCACGACCAACATGTTTACTTACAGCGGTTTATCAATTCACCT 2195  
DB 203 CCTAAGAGTGTGATTCACGACCAACATGTTTACTTACAGCGGTTTATCAATTCACCT 144  
QY 2196 TGAGTAACAACATGTTTCACTGTCAATTTATAGTGTGCGGACCAAGTTGGTTGAGT 2255  
DB 143 TGAGTAACAACATGTTTCACTGTCAATTTATAGTGTGCGGACCAAGTTGGTTGAGT 84  
QY 2256 ATACTGTTTGCATATAGACTGGGTGCTAAATTTCTGGTGAAGCATTTTATACCA 2315  
DB 83 ATACTGTTTGCATATAGACTGGGTGCTAAATTTCTGGTGAAGCATTTTATACCA 24  
QY 2316 TTGTAGGCTTTTAACTCTTGA 2338  
DB 23 TTGTAGGCTTTTAACTCTTGA 1

RESULT 3  
US-09-988-863a-4  
Sequence 4, Application US/09988863A  
Patent No. US20020123427A1  
GENERAL INFORMATION:  
APPLICANT: Bayer AG  
TITLE OF INVENTION: Plant phosphomevalonate kinases  
FILE REFERENCE: Le A 35 018  
CURRENT APPLICATION NUMBER: US/09/988,863A  
PRIORITY FILING DATE: 2001-11-21  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 728  
TYPE: DNA  
ORGANISM: Gossypium hirsutum  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 715  
OTHER INFORMATION: n can be any nucleotide  
US-09-988-863a-4

Query Match 17.8%; Score 425.4; DB 10; Length 728;  
Best Local Similarity 76.1%; Pred. No. 6,8e-116;  
Matches 554; Conservative 0; Mismatches 162; Indels 12; Gaps 2;

QY 784 CGGTTTACGCGATTTGTGAAGCCATCAACGAGAAAGTAAAGCTGGAAGTGGCGATG 843  
DB 1 CGGTTTACGCGATTTGTGAAGCCATCAACGAGAAAGTAAAGCTGGAAGTGGCGATG 60  
QY 844 AATGAGACAGATGTCAAAATTAATCAATCACCAGCTCTGAGAGAAAGCATGTATAACTG 903  
DB 61 TCTTGACCGAGATGTCAAGTCAATCAATCTCTGAGAGAAAGCATGTATAACTG 120  
QY 904 TCATGTAATCATTTGACCTGCTGCTGTGTCGCAAGTATCAAGAAACCCCTTTGTA 963  
DB 121 TCTGGAACCATTTTAACACTTCAAGTGTATCTTCAAGTAAAGTAAAGAAACCTTTGTA 180  
QY 964 GAGCATGCATACAGTATCTATAGCTGCTGCTCATTTGGCA---ACGAGAAAGCAAA 1020  
DB 181 GAAATGCTATTCATATATCTATAGCACTGCATGCAATGCAATTTGCAAGATTAAGAA 240  
QY 1021 GAATCATTCACAACCTTATATGCAAGCTTGTATATACATATTTAGCTCAATGAC 1080  
DB 241 GAGGCTTTGATTAATCACTTCTTACAGGCTTGTATATACATTTAGCTTCAATGAC 300  
QY 1081 TTTTACTATTCGGAACCATAGATAGATGAGCTGGCTTCCATTTGACACCAAGTGC 1140  
DB 301 TTTTACTATTCGGAACCATAGATAGATGAGCTGGCTTCCATTTGACACCAAGTGC 360  
QY 1141 GGTACCTTGCACCGTTTGCATCAATCAATTCATGCTGCGAGTCAAAAGTGTAT 1200  
DB 361 GGTACCTTGCACCGTTTGCATCAATCAATTCATGCTGCGAGTCAAAAGTGTAT 420  
QY 1201 TCCAGGCTGAGTGAACCAAACTGGCTGAGTCTTGTGCAAGATGACACAGCTG 1260  
DB 421 TCCAGGCTGAGTGAACCAAACTGGCTGAGTCTTGTGCAAGATGACACAGCTG 480  
QY 1261 GTTGACGCTCTTCAATATCTTGGAGTGTGACATCTGATTCATGATGTAAG--- 1315  
DB 481 GTTGACGCTCTTCAATATCTTGGAGTGTGACATCTGATTCATGATGTAAG--- 540  
QY 1316 ---AAGGAAGTTGGCTGCTCTATCTAGATGTTATTCATATGATACCAAAAGCT 1371  
DB 541 CACCAAGAAATTAAGATTCACAGATTCGATATTTGTCATATGATACCTAAAGTGC 600  
QY 1372 CATGCTTGCACAAAGGAGTGGAGTGGTGTGATGATGATGATGATGATGATGATGAT 1431  
DB 601 CACTGATTTGCCCAAGGATTAAGTGGAGTGGCTTGAATGATGATGATGATGATGATG 660  
QY 1432 AGTCAGCGTTATGTTGCTCTCTCCAGAGTCTTGTCTGATTTGCTCAGGTTGACATACA 1491  
DB 1432 AGTCAGCGTTATGTTGCTCTCTCCAGAGTCTTGTCTGATTTGCTCAGGTTGACATACA 1491

Db 661 AGTACGGTATGTTCTTTTACACCAAAAGTCTTCTGCTGCTGACGGCTGCANTGAA 720  
Qy 1492 GGTCTGCC 1499  
Db 721 GGATGCC 728

RESULT 4  
US-09-988-863a-3  
; Sequence 3, Application US/09988863A  
; Patent No. US20020123427A1  
; GENERAL INFORMATION:

; APPLICANT: Bayer AG  
; TITLE OF INVENTION: Plant phosphomevalonate kinases  
; FILE REFERENCE: Le A 35 018  
; CURRENT APPLICATION NUMBER: US/09/988, 863A  
; CURRENT FILING DATE: 2001-11-21  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 611  
; TYPE: DNA  
; ORGANISM: Medicago truncatula  
US-09-988-863a-3

Query Match 12.5%; Score 299.2; DB 10; Length 611;  
Best Local Similarity 71.7%; Pred. No. 2.2e-78;  
Matches 407; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

Qy 691 GTTGTGCTTCTGCTGCTGGAAGTTTGATGACTGAGGCTGCTGCTGCTGAGAG 750  
Db 44 GTTGTGCTTCTGCTGCTGGAAGTTTGATGACTGAGGCTGCTGCTGCTGAGAG 103  
Qy 751 CCAATGAGGCTTGTGTGATGACAATGCAAGGTTTACGGATTTGAGCAATC 810  
Db 104 CCTATGCTGAGCTTGTGTGATGACAATGCAAGGTTTACGGATTTGAGCAATC 163  
Qy 811 AACCAAGATGCAAGCTGGAAGTTGGCATGGAATGGAAGATGTCATTAATCA 870  
Db 164 TATCTCAAACTAAACGATGTTGGCTTGGCTGAGATGTCAGATTAAACATC 223  
Qy 871 CCACAGCTGAGAGAAAGATGATTAATGCTGATGATGATGATGATGATGATG 930  
Db 224 CCTAAGCTGAGAGAAAGATGATTAATGCTGATGATGATGATGATGATGATG 283  
Qy 931 GTGTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 990  
Db 284 GTTCTCAAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 343  
Qy 991 GCTGCTGATTTGGCAAC---CGAAGAGCAAAAGATGATGATGATGATGATG 1047  
Db 344 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 403  
Qy 1048 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1107  
Db 404 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 463  
Qy 1108 TCGGCTGAGCTTCAATGACCAAGATGCTGGTGAACCTTGGACCTTTGCAATC 1167  
Db 464 AGACAGGACTCCCTTGACATCAGAAATCATTTGGGCAACCTTTGGCCCTCA 523  
Qy 1168 ACATTCATGCTGCGGAGTCAATGCTGATTTCCAAAGCTGAAAGTGAAGTGGC 1227  
Db 524 TCTTCAATGCTGCGGAGTCAATGCTGATTTCCAAAGCTGAAAGTGAAGTGGC 583  
Qy 1228 TTAGGTTCTTCTGAGCAATGACCAAG 1255  
Db 584 TTGGGCTATCTGACGATGACCAAG 611

RESULT 5  
US-09-909-745-21  
; Sequence 21, Application US/09909745

; Patent No. US20020119546A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Farnold, Omolayo O.  
; TITLE OF INVENTION: Squalene Synthase Enzymes  
; FILE REFERENCE: BB112 US CIP  
; CURRENT APPLICATION NUMBER: US/09/909, 745  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 21  
; LENGTH: 757  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
US-09-909-745-21

Query Match 9.6%; Score 229.2; DB 10; Length 757;  
Best Local Similarity 61.3%; Pred. No. 1.8e-57;  
Matches 369; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

Qy 1601 GAAGTGTGATCTCTCCACACATCATGATGATGATGATGATGATGATGATG 1660  
Db 9 GTATCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 68  
Qy 1661 ATCCAGAGAGGACGAGAAATGCGGAGATTTGTGATGATGATGATGATGATG 1720  
Db 69 ACCCTCAAAATTCAGAGAGATGATGATGATGATGATGATGATGATGATGATG 128  
Qy 1721 CTAAGCTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1780  
Db 129 ACCAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 188  
Qy 1781 TTAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1840  
Db 189 TAGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 248  
Qy 1841 AGAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1900  
Db 249 AAGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308  
Qy 1901 TTAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1960  
Db 309 ATATGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 368  
Qy 1961 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2020  
Db 369 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428  
Qy 2021 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2080  
Db 429 GCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488  
Qy 2081 GGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2140  
Db 489 GGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 548  
Qy 2141 AAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2200  
Db 549 AAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 608  
Qy 2201 AA 2202  
Db 609 GA 610

RESULT 6  
US-09-909-745-19  
; Sequence 19, Application US/09909745  
; Patent No. US20020119546A1



```

: :
: : NUMBER OF SEQ ID NOS: 875
: :
: : SOFTWARE: FastSeq for Windows Version 4.0
: :
: : SEQ ID NO 273
: :
: : LENGTH: 2004
: :
: : TYPE: DNA
: :
: : ORGANISM: Arabidopsis thaliana
US-03-887-576-273

```

Query Match	5.1%;	Score 123.2;	DB 10;	Length 2004;
Best Local Similarity	-71.28;	Pred. No. 1.2e-25;		
Matches 195;	Conservative	0;	Mismatches 68;	Indels 11;
				Gaps 2

OY	271	TAGTGGTTTATTTTCACAGAGTTTTGGCTTCCTTTATATGGGACACAGACGTCGA	330
Db	512	TAGTGTTTTTTGGGCATTCACGCTGTGGTTCCTCTCGACGGCGGACG-TTTTTCAGGC	455
OY	331	GTTCCTGTCAAAAGCTTACAGATCCGACTCGAGCGCTGACAGTAAAGAAAGACAGCG	390
Db	453	GATTGTATCTCTCGATATCATCTGTCCTCATTTAGTCAAAAGGAACAAACAGGCA	394
OY	391	ATTGTCGTC-----TAGATGAGCGGCAAGCGTGTGCATCCGTCGATCGACGG	444
Db	333	ATTGTGTGGCATTCGCTCTGCATCAACGGCCATGTGTGTGCATCCGCTCAATCGACGG	333
OY	441	AGAAATACGTTTCGATCCGGTTTCGATCCAAATCGAGAGTTTGAAGATCTAAATCGGAAA	500
Db	333	CGATAGTGTGTCGATCGATCGTCATTCATCCAAATCGAGAGTTTGAAGTTCGAAATCGAGA	274
OY	501	TTGCATTAATACATCATCCCAATCTCTTCRAG	534
Db	273	ATTATATGATCTCATCTCCACATCTCTCTTCAG	240

RESULT 9  
 : US-09-909-745-15/c  
 : Sequence 15, Application US/09909745  
 : Patent No. US20020119546A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Falco, S. Carl  
 : APPLICANT: Famodu, Omolayo O.  
 : TITLE OF INVENTION: Squelene Synthesis Enzymes  
 : FILE REFERENCE: B0112 US CIP  
 :  
 : CURRENT APPLICATION NUMBER: US/09/909, 745  
 : CURRENT FILING DATE: 2001-07-20  
 :  
 : PRIOR APPLICATION NUMBER: 60/107,241  
 : PRIOR FILING DATE: 1998-11-05  
 :  
 : PRIOR APPLICATION NUMBER: 60/107,241  
 : PRIOR FILING DATE: 1998-11-05  
 :  
 : NUMBER OF SEQ ID NOS: 26  
 :  
 : SOFTWARE: Microsoft Office 97  
 :  
 : SEQ ID NO 15  
 :  
 : LENGTH: 249  
 :  
 : TYPE: DNA  
 :  
 : ORGANISM: Zea mays  
 :  
 : US-09-909-745-15

Query Match	4.5%	Score 107.2;	DB 10;	Length 249;
Best Local Similarity	64.5%	Pred. No. 1.7e-21;		
Matches 160; Conservative	0;	Mismatches 88;	Indels 0;	Gaps 0

QY	1789	TGTTGTTGCTTACTTCTTGAAAAGGTGGGTGTTACATGTCATCGAACCAATCAAGAACGC	1844
Db	249	TGTAGTGTCTTCACATATGCGAAATGGGGCAGAGTGGCTACTACCAACATCAACAATAATTA	190
QY	1849	ATTATTAAGAACCTCTAGAGGCGACAGAAAGCTATGTGAGATCGAATTTCTTATGCGT	1908
Db	189	ATTATTAAGATCATGTTGGCTGCGAAGGATGCTTGGCTTGACATTAAGGCTTTACATGCGA	130
QY	1909	CAGATGGGTGAGCGGCTAGCGTTCGCGATAGACCTGAATCTCAAACTCAACTTTTGGAT	1968
Db	129	GAGATGGGCAATAGACAGCTGGTGTTCCAATTAGACCAGATTCACAAACAGCGCTACTAGAT	70
QY	1969	TCCTCAATGAGTGTCTGAAGGATTTCTACTTCTGTTGTTCTCGAGACCTGGTGATTTGAT	2028

Db	QY	2029	2036	10
69	GCACCATATGAAATGAGAGCGTCTTTACTGCGTGGAGTTCCGAGCGTGGTGGCTTCGAT	11	11	11
9	GCACTGTT	2	2	2

```

RESULT 10
US-09-909-745-17
Sequence 17, Application US/09909745
Patent No. US20020119546A1
GENERAL INFORMATION:
APPLICANT: Falco, S. Carl
APPLICANT: Famodu, Omolayo O.
TITLE OF INVENTION: Squalene Synthe
FILE REFERENCE: Bb1112 US CIP
CURRENT APPLICATION NUMBER: US/09/9
PRIOR FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: 60/107,24
PRIOR FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 60/107,24
PRIOR FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Microsoft Office 97
SEQ ID NO 17
LENGTH: 249
TYPE: DNA
ORGANISM: zea mays
US-09-909-745-17

```

Query Match	4.58;	Score 107.2;	DB 10;	Length 249;
Best Local Similarity	64.58;	Pred. No. 1.7e-21;		
Matches 160;	Conservative 0;	Mismatches 88;	Indels 0;	Gaps 0

[illegible]

```

RESULT 11
US-09-938-842A-4887
: Sequence 4887, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kireps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPI300-3
: CURRENT APPLICATION NUMBER: US/09/938,842A
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647

```



Query Match 2.4%; Score 57.6; DB 9; Length 7681;  
Best Local Similarity 50.7%; Pred. No. 1.1e-05;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Search completed: May 3, 2003, 21:10:48  
Job time : 234 secs

QY 1337 ATCTAGATGTTATCCATATGATACACAAAGCTCTCATGTCTTCGACAGGAGGCTCG 1396  
DB 585 ATAGAGAGATTATTCATTAATTTACCAAGTTGCTCATGTTCACAGCTCAGGTAATAATG 644  
QY 1397 GAAGTGGTTTGAATGAGCTGCTGCTGCTATGAGAGTACAGCTTATGTTGCTCTCTC 1456  
DB 645 GAACGGGTTTGAATGAGCTGCTGCTGCTATGAGAGTACAGCTTATGTTGCTCTCTC 704  
QY 1457 CAGAGCTTGTCTATTTGCTCAGGTTGCTCAGTACAGCTGCTGCTTAAATGAGTTATG 1516  
DB 705 CCGCATTAATCTTAATTTGCGACATATGGAAGTGTACTTACGCGAGTAACCTGGCGC 764  
QY 1517 GTACATTTTGAAGGAAATGGACATTAAGAGACTGAGTTCTTTTACCACTGCA 1576  
DB 765 ATTTGTTGATGAGAGAGACTGAGTATTTAGATTAAGTAAGTAAATTAACCTTCGGGAT 824  
QY 1577 TGAATCTTTCTTCTGAGAGACCTGGAAGTGT 1608  
DB 825 TAACCTTATGATGGCGCATATTTAAGATGCT 856

## RESULT 15

US-09-918-740-58  
; Sequence 58, Application US/09918740  
; Publication No. US20030033626A1  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Frederick  
; APPLICANT: Kuehnle, Adelheid  
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways  
; FILE REFERENCE: KAS-103XCI  
; CURRENT APPLICATION NUMBER: US/09/918,740  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/221,703  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 58  
; LENGTH: 7693  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Operon containing A. thaliana and S. cerevisiae DNA  
US-09-918-740-58

Query Match 2.4%; Score 57.6; DB 9; Length 7693;  
Best Local Similarity 50.7%; Pred. No. 1.1e-05;  
Matches 138; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1337 ATCTAGATGTTATCCATATGATACACAAAGCTCTCATGTCTTCGACAGGAGGCTCG 1396  
DB 4338 ATAGAGAGATTATTCATTAATTTACCAAGTTGCTCATGTTCACAGCTCAGGTAATAATG 4397  
QY 1397 GAAGTGGTTTGAATGAGCTGCTGCTGCTATGAGAGTACAGCTTATGTTGCTCTCTC 1456  
DB 4398 GAACGGGTTTGAATGAGCTGCTGCTGCTATGAGAGTACAGCTTATGTTGCTCTCTC 1456  
QY 1457 CAGAGCTTGTCTATTTGCTCAGGTTGCTCAGTACAGCTGCTGCTTAAATGAGTTATG 1516  
DB 4458 CCGCATTAATCTTAATTTGCGACATATGGAAGTGTACTTACGCGAGTAACCTGGCGC 4517  
QY 1517 GTACATTTTGAAGGAAATGGACATTAAGAGACTGAGTTCTTTTACCACTGCA 1576  
DB 4518 ATTTGTTGATGAGAGAGACTGAGTATTTAGATTAAGTAAGTAAATTAACCTTCGGGAT 4577  
QY 1577 TGAATCTTTCTTCTGAGAGACCTGGAAGTGT 1608  
DB 4578 TAACCTTATGATGGCGCATATTTAAGATGCT 4609

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 17:57:36 / Search time 66 Seconds  
(without alignments)  
11133.296 Million cell updates/sec

Title: US-09-988-863A-1

Perfect score: 2396  
Sequence: 1 gtcgacgcacgcgtccgycg.....ttctcaaaaaaaaaaaaaa 2396

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTCUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
C	1	44.6	1.9	7218	1	US-08-232-463-14
	2	37	1.5	65042	4	US-09-784-316-3
	3	35.2	1.5	18627	4	US-08-961-527-113
C	4	35.2	1.5	168575	4	US-09-426-290-1
	5	34.6	1.4	169998	4	US-09-676-610B-24
	6	34.4	1.4	1896	4	US-09-552-332-5
C	7	34.2	1.4	1395	1	US-07-991-867B-25
	8	34.2	1.4	1395	1	US-08-107-755A-25
C	9	34.2	1.4	1395	2	US-08-544-332-25
	10	34.2	1.4	1395	2	US-09-370-861A-25
C	11	34.2	1.4	6768	1	US-08-107-755A-1
	12	34.2	1.4	8457	1	US-07-991-867B-1
C	13	34.2	1.4	8457	2	US-08-544-332-1
	14	34.2	1.4	8457	4	US-09-370-861A-1
C	15	33.6	1.4	1062	4	US-09-253-316-3
	16	33.6	1.4	1539	4	US-09-364-230-29
	17	33.6	1.4	2287	4	US-08-845-258-8
C	18	33.6	1.4	2287	4	US-08-990-571-8
	19	33.6	1.4	2287	4	US-08-722-142A-8
	20	33.6	1.4	2287	4	US-09-528-784A-8
	21	33.2	1.4	927	2	US-07-690-192-1
	22	33.2	1.4	1215	4	US-09-134-001C-277
	23	33.2	1.4	1485	4	US-09-134-001C-2187
	24	33.2	1.4	1753	4	US-09-149-476-56
	25	33.2	1.4	1939	4	US-09-357-251-25
	26	33	1.4	2200	1	US-08-272-255-21
	27	33	1.4	2200	5	PCT-US95-08565-21

C	28	33	1.4	3418	4	US-09-590-101A-7	Sequence 7, Appli
	29	33	1.4	3678	4	US-09-590-101A-9	Sequence 9, Appli
C	30	33	1.4	7218	4	US-08-232-463-14	Sequence 14, Appli
	31	32.8	1.4	4002	3	US-09-356-952-9	Sequence 9, Appli
	32	32.6	1.4	750	4	US-09-134-001C-1073	Sequence 1073, Ap
	33	32.4	1.4	1364	4	US-08-265-087-3	Sequence 3, Appli
	34	32.4	1.4	1364	1	US-08-621-493-3	Sequence 3, Appli
	35	32.4	1.4	1364	2	US-08-965-688-3	Sequence 3, Appli
	36	32.4	1.4	1364	4	US-09-260-173-3	Sequence 3, Appli
	37	32.4	1.4	162450	4	US-09-345-882-1	Sequence 1, Appli
	38	32.2	1.3	1146	2	US-08-458-109-41	Sequence 41, Appli
	39	32.2	1.3	10395	1	US-08-245-809-3	Sequence 3, Appli
	40	32.2	1.3	10396	1	US-08-245-809-5	Sequence 5, Appli
	41	32.2	1.3	10798	1	US-08-107-748-2	Sequence 2, Appli
	42	32.2	1.3	10798	5	PCT-US92-01385-2	Sequence 2, Appli
	43	32.2	1.3	10965	1	US-08-107-748-4	Sequence 4, Appli
	44	32.2	1.3	10965	5	PCT-US92-01385-4	Sequence 4, Appli
	45	32	1.3	1246	4	US-09-302-769-22	Sequence 22, Appli

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F15  
US-08-232-463-14  
Query Match 1.9%; Score 44.6; DB 1; Length 7218;

Query Match	1.5%:	Score 37:	DB 4:	Length 65042:
Best Local Similarity	53.0%:	Pred. No. 3.3:		
Matches 79:	Conservative 0:	Mismatches 70:	Indels 0:	Gaps 0:
QY 2164	TGTATTACTCAGGCGTTTCATCAATTCACCTTGAGTAACACAACTGTTTCAGTGCCA	2223		
Db 44535	TTTTTTCATTCTATTTCTCTGTATTCAGATGSGGTATTTACATTTGTTCAATTTTCCA	44594		
QY 2224	ATTATTAGGCGCGTCACCAAGTTGCGTTGAGTATACGTGTTTGCATPANAAGACTTGGGTC	2283		
Db 44595	GTTCAATTGATTATTTCCCTCTGCCCCCTGCATCTGTTGTTGAGCCTACTACTGAGCTTT	44654		
QY 2284	TAAATTCTTGCGTAAAGCATTTTATAC	2312		
Db 44655	TTATTTTGTTATATGTAATTTTATTAATTC	44683		

```

/ ORGANISM: HOMO SAPIENS
:
: FEATURE:
:

```



NAME/KEY: CDS  
LOCATION: (21181)...(21403)  
NAME/KEY: CDS  
LOCATION: (95252)...(95430)  
NAME/KEY: CDS  
LOCATION: (101753)...(101996)  
NAME/KEY: CDS  
LOCATION: (110324)...(110439)  
NAME/KEY: CDS  
LOCATION: (124058)...(124278)  
NAME/KEY: CDS  
LOCATION: (127009)...(127130)  
NAME/KEY: CDS  
LOCATION: (128910)...(129139)  
US-09-426-290-1

Query Match 1.5%; Score 35.2; DB 4; Length 168575;  
Best Local Similarity 52.0%; Pred. No. 21;  
Matches 79; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 2245 TTGGTGTAGTATCTGTTTGCATATAGACTGGGTCTAAATTTCTGGTGAAGCAT 2304  
DB 79162 TTGATATATTAACCTGATCTGCATTCAGACATTTGATTAACATTTCTAGTTTAGGGCT 79103  
QY 2305 TTTTATACCATTTGTAGGCTTTTAACTCTTGAAACTGGGGAATAAATTAAGT 2364  
DB 79102 CACACTGAGTAAAGTAAAGATGATATCTTTAGACACATGTCGAGAAAGTTAGATTCA 79043  
QY 2365 TGATTTCAATCTCTCAAAAAA 2396  
DB 79042 TAAATTCGACTCTATTCAGAAAGAACACACA 79011

## RESULT 5

US-09-676-610B-24  
Sequence 24, Application US/09676610B  
Patent No. 6444465  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
APPLICANT: Jacqueline Wyatt  
APPLICANT: Susan M. Freiler  
TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION  
FILE REFERENCE: RTS-0138  
CURRENT APPLICATION NUMBER: US/09/676,610B  
CURRENT FILING DATE: 2000-09-29  
NUMBER OF SEQ ID NOS: 182  
SEQ ID NO 24  
LENGTH: 169998  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: (1208)...(1472)  
NAME/KEY: intron  
LOCATION: (1473)...(124390)  
NAME/KEY: exon  
LOCATION: (124391)...(124544)  
NAME/KEY: intron  
LOCATION: (124545)...(125409)  
NAME/KEY: exon  
LOCATION: (125410)...(125595)  
NAME/KEY: intron  
LOCATION: (125596)...(128711)  
NAME/KEY: exon  
LOCATION: (128712)...(128848)  
NAME/KEY: intron  
LOCATION: (128849)...(133400)  
NAME/KEY: exon  
LOCATION: (133401)...(133469)  
NAME/KEY: intron  
LOCATION: (133470)...(134652)  
NAME/KEY: exon  
LOCATION: (134653)...(134773)

NAME/KEY: intron  
LOCATION: (134774)...(136116)  
NAME/KEY: exon  
LOCATION: (136117)...(136261)  
NAME/KEY: intron  
LOCATION: (136262)...(137936)  
NAME/KEY: exon  
LOCATION: (137937)...(138053)  
NAME/KEY: intron  
LOCATION: (138054)...(138637)  
NAME/KEY: exon  
LOCATION: (138638)...(138766)  
NAME/KEY: intron  
LOCATION: (138767)...(138864)  
NAME/KEY: exon  
LOCATION: (138865)...(138940)  
NAME/KEY: intron  
LOCATION: (138941)...(139765)  
NAME/KEY: exon  
LOCATION: (139766)...(139860)  
NAME/KEY: intron  
LOCATION: (139861)...(142245)  
NAME/KEY: exon  
LOCATION: (142246)...(142445)  
NAME/KEY: intron  
LOCATION: (142446)...(143605)  
NAME/KEY: exon  
LOCATION: (143606)...(143738)  
NAME/KEY: intron  
LOCATION: (143739)...(145838)  
NAME/KEY: exon  
LOCATION: (145839)...(145931)  
NAME/KEY: intron  
LOCATION: (145932)...(147385)  
NAME/KEY: exon  
LOCATION: (147386)...(147544)  
NAME/KEY: intron  
LOCATION: (147545)...(153274)  
NAME/KEY: exon  
LOCATION: (153275)...(153321)  
NAME/KEY: intron  
LOCATION: (153322)...(155088)  
NAME/KEY: exon  
LOCATION: (155089)...(155231)  
NAME/KEY: intron  
LOCATION: (155232)...(156025)  
NAME/KEY: exon  
LOCATION: (156026)...(156151)  
NAME/KEY: intron  
LOCATION: (156152)...(156826)  
NAME/KEY: exon  
LOCATION: (156827)...(156928)  
NAME/KEY: intron  
LOCATION: (156929)...(163399)  
NAME/KEY: exon  
LOCATION: (163400)...(163586)  
US-09-676-610B-24

Query Match 1.4%; Score 34.6; DB 4; Length 169998;  
Best Local Similarity 50.3%; Pred. No. 33;  
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 843 GAAATGACAGATGTCATTAATTAACCCACAGCTGCGAGAGAAAGCATGTAACT 902  
DB 51241 GAAATTCACCTAATGCAATTAACCTAGCAAAATGCTATATGTTATTTTCTTAAT 51300  
QY 903 GTCACTGATCAATTTGACTCTTCAGTCTGTGTGCAAGTGAATTCAGAACCCCTTGT 962  
DB 51301 TCAGATTAATTAAGACAAATTCATTCCTGAAATGCTGTGTCATGTAAGAAAGAAATTTAT 51360  
QY 963 AGAGCATGCGATACAGTATGCTATAGCTGCTCTCATATTTGGACACCGAG 1011  
DB 51361 CGAGGTGGCCCTTGAGTGCCAAACAGCCGTGCTCAGCTGCAAAATGAG 51409

RESULT 6  
US-09-552-322-5  
; Sequence 5, Application US/09552322  
; Patent No. 6436642  
; GENERAL INFORMATION:  
; APPLICANT: Gould-Rothberg  
; APPLICANT: Rastelli  
; TITLE OF INVENTION: METHOD OF CLASSIFYING A THYROID CARCINOMA USING  
; TITLE OF INVENTION: DIFFERENTIAL GENE EXPRESSION  
; FILE REFERENCE: 15966-548  
; CURRENT APPLICATION NUMBER: US/09/552,322  
; CURRENT FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: 60/130,123  
; PRIOR FILING DATE: 1999-04-20  
; PRIOR APPLICATION NUMBER: 60/193,203  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1896  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Assembled  
; OTHER INFORMATION: using sequences from AC024267  
; NAME/KEY: misc.feature  
; LOCATION: (25)..(1291)  
; OTHER INFORMATION: wherein any n is a, g, c, or t.  
US-09-552-322-5

Query Match 1.4%; Score 34.4; DB 4; Length 1896;  
Best Local Similarity 59.0%; Pred. No. 2.5;  
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 2297 GTAGCATTTTATACCATGTAGCTTACTTACTCTGGAACACTTGGCGGAAATAA 2356  
Db 1796 GTACCAATGACATGATGATGATGATGATGATGATGATGATGATGATGATG 1855  
Qy 2357 AATAAGTGTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 2396  
Db 1856 TATATATAAACAATAATCTCTCAAAAAAAAAAAAAAAAAAAAAA 1895

RESULT 7  
US-07-991-867B-25/C  
; Sequence 25, Application US/07991867B  
; Patent No. 5476781  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Hall, Richard L.  
; APPLICANT: Gruidl, Michael E.  
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/991,867B  
; FILING DATE: 12-DEC-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO 92/14818

FILING DATE: 12-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,685  
; FILING DATE: 30-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,584  
; FILING DATE: 19-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF114.C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-07-991-867B-25

Query Match 1.4%; Score 34.2; DB 1; Length 1395;  
Best Local Similarity 56.8%; Pred. No. 2.3;  
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 1936 ATAGAGCCTGATCTCAACATCACTTTTGTGATCTCAATGAGTCTGAGAGTTCTA 1995  
Db 291 ATAAATCTAATTTGATTTAAATAATTAATTTATTTATTTATTTATTTATTTATTT 232  
Qy 1996 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2046  
Db 231 TTGCTAGTAGTATAGAGTGATTTATTTATTTATTTATTTATTTATTTATTTATTT 181

RESULT 8  
US-08-107-755A-25/C  
; Sequence 25, Application US/08107755A  
; Patent No. 5721352  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Hall, Richard L.  
; APPLICANT: Gruidl, Michael E.  
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: U.S.A.  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/107,755A  
; FILING DATE: 19-AUG-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,658  
; FILING DATE: 30-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,584  
; FILING DATE: 19-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF114.C2  
; TELECOMMUNICATION INFORMATION:

; INFORMATION FOR SEQ ID NO: 25:

5721352e1 Entomopoxvirus Expression System



```

Yy 1936 ATGAGCGCTAAATCTGAACTCACTTGGAGTTCACATGAGCGGAGAGACTTCTA 1993
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 ATTAATATCTAAATTTGATTTAAATAAATATATATTTCTATGATTTATGGCATTTA 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1996 CTTCCTGCTGTTCTCTGAGCTGTGGATTTGATGACCAATTTGCATCACT 2046
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 TTGCTGATGATGTTATAGAGGTGCATTTATTTTCTAGATATGATTAAT 245

RESULT 13
: Sequence 1, Application US/08544332
: Patent No. 5935777
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Gruidl, Michael E.
: TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
: NUMBER OF SEQUENCES: 77
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gerard H. Bencen
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/544,332
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/991,867
: FILING DATE: 07-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/107,755
: FILING DATE: 19-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO 92/14818
: FILING DATE: 12-FEB-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/827,685
: FILING DATE: 30-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/657,584
: FILING DATE: 19-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Bencen, Gerard H.
: REGISTRATION NUMBER: 35,746
: REFERENCE/DOCKET NUMBER: UF114.C4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8457 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Amsacta moorei entomopoxvirus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: complement (65..1459)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1474..2151

```

```

?       FEATURE: CDS
?       LOCATION: complement (2239..2475)
?
?       FEATURE:
?       NAME/KEY: CDS
?       LOCATION: 2502..2987
?       FEATURE:
?       NAME/KEY: CDS
?       LOCATION: 3080..6091
?       FEATURE:
?       NAME/KEY: CDS
?       LOCATION: complement (6277..6768)
?
US-08-544-332-1

Query Match      1.4%: Score 34.2; DB 2; Length 8457;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1936 ATAGAGCCGTGATCTCAACACTCACCATTGGATTTGTACAAATGAGTGTGAAGAGTTCTA 1995
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 ATAAAATCTAATATTGGAATTTAAAAAAATTAATATTCTAGTAGAGTTTATTGGTGATTTA 296

QY 1996 CTTCGTGTGTTCCCGAGAGCTGGTGATTTGCATGCCATATTTCGCAATCACT 2046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 TTGCCTAGTAGTTATAGAGGTGCATTAATTATTCTAGATAGATTAAT 245


RESULT 14
US-09-370-861A-1/c
Sequence 1, Application US/09370861A
Patent No. 6410221
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OR INVENTION: No. 6410221el Entomopoxvirus. Expression System
FILE REFERENCE: UPI14.C4.DI
CURRENT APPLICATION NUMBER: US/09/370.861A
CURRENT FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: US 07/991,867
PRIOR FILING DATE: 1992-12-07
PRIOR APPLICATION NUMBER: US 08/107,755
PRIOR FILING DATE: 1993-08-19
PRIOR APPLICATION NUMBER: WO 92/14818
PRIOR FILING DATE: 1992-02-12
PRIOR APPLICATION NUMBER: US 07/827,685
PRIOR FILING DATE: 1992-01-30
PRIOR APPLICATION NUMBER: US 07/657,584
PRIOR FILING DATE: 1991-02-19
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 8457
TYPE: DNA
ORGANISM: Amsacta moorei entomopoxvirus
US-09-370-861A-1

Query Match      1.4%: Score 34.2; DB 4; Length 8457;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1936 ATAGAGCCGTGATCTCAACACTCACCATTGGATTTGTACAAATGAGTGTGAAGAGTTCTA 1995
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 ATAAAATCTAATATTGGAATTTAAAAAAATTAATATTCTAGTAGAGTTTATTGGTGATTTA 296

QY 1996 CTTCGTGTGTTCCCGAGAGCTGGTGATTTGCATGCCATATTTCGCAATCACT 2046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 TTGCCTAGTAGTTATAGAGGTGCATTAATTATTCTAGATAGATTAAT 245


RESULT 15
US-09-253-316-3/c
Sequence 3, Application US/09253316
```

Patent No. 6395890  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Jaspers, Stephen R.  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR HOMOLOGS  
FILE REFERENCE: 97-75  
CURRENT APPLICATION NUMBER: US/09/253,316  
CURRENT FILING DATE: 1999-02-19  
EARLIER APPLICATION NUMBER: US 60/075,300  
EARLIER FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 1062  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Degenerate sequence  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1062)  
OTHER INFORMATION: n = A,T,C or G  
US-09-253-316-3

Query Match 1.48; Score 33.6; DB 4; Length 1062;  
Best Local Similarity 28.8%; Pred. No. 3;  
Matches 78; Conservative 49; Mismatches 143; Indels 1; Gaps 1;  
QY 1999 GCTGCTGTCCTGAGCTGGTGGATTTGATGCCATATTTGCATTCACATTAGGGGATTC 2058  
DB 830 GCYTTSMNARYTGRAANGNGSYTGRCANGYTTTNCCTTNGDATTITDANGYTTT 771  
QY 2059 GGCACCAACTGACCCAGCGCATGGAGTTGCGACAAATGTTTGGCCTTGTGTGAGAGAA 2118  
DB 770 ARDATRTNSMRYRCANGSYTGDAITRARCANAARNCKYTTTCYTNCKCAT-YTCRCA 712  
QY 2119 GATCCACATGGCCTTGGCCAGAAAGTGGATGCCAGAACACATGTATTACTCAGGC 2178  
DB 711 RTNSMRYTYTCRTTNGTNCNCRTTNSWDATNCCCATNCCRCANGTNCNSMRCANGG 652  
QY 2179 GTTTCATCATTCACCTTGAGTAACAACATGTTTCAGTGCCAAATTATTAGTGGTC 2238  
DB 651 NGTCCAYTTNGTNGCYTGNACNARCAVTTYTTTCCADATNARNGNARRTTNCKRTA 592  
QY 2239 ACCAAGTTGGTTGAGTATACGTGTTGCAT 2269  
DB 591 NGCNGCATNGTYTTRTANSWNGTNSMNAV 561

Search completed: May 3, 2003, 20:11:34  
Job time : 1035 secs

GenCore version 5.1.4.p5.4578.  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:53:20 ; Search time 88 Seconds  
(without alignments)  
1182.429 Million cell updates/sec

Title: US-09-988-863A-2  
Perfect score: 2581  
Sequence: 1 MAVVASAPKVKLTGGLV.....ESGDPRTTCITSGVSIHLE 505

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_mammal:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_protist:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacteriophage:\*  
18: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2581	100.0	505	10 Q9C6T1	Q9C6T1 arabidopsis
2	1930.5	74.8	503	10 Q944G1	Q944G1 hevea bras
3	493	19.1	426	3 Q9U788	Q9U788 schizosacch
4	235	9.1	314	17 Q973B5	Q973B5 sulfolobus
5	230	8.9	323	17 Q97UL6	Q97UL6 sulfolobus
6	178.5	6.9	368	2 Q9FDE7	Q9FDE7 enterococcu
7	158.5	6.1	294	2 Q9K31	Q9K31 lactobacilli
8	155	6.0	374	2 Q9KWC3	Q9KWC3 streptomyce
9	151.5	5.9	360	16 Q92FUI	Q92FUI listeria in
10	149.5	5.8	359	16 Q8YAV1	Q8YAV1 listeria in
11	147.5	5.7	361	2 Q9FDE2	Q9FDE2 enterococcu
12	142.5	5.5	358	2 Q9FDE2	Q9FDE2 enterococcu
13	136.5	5.3	358	2 Q9FDE3	Q9FDE3 staphylococ
14	136.5	5.3	360	2 Q9KWF7	Q9KWF7 staphylococ
15	134.5	5.2	358	16 Q99W20	Q99W20 kitasatospo
16	130.5	5.1	358	2 Q9FDF7	Q9FDF7 staphylococ

17	130.5	5.1	434	11 Q9D307	Q9D307 mus musculu
18	130	5.0	816	5 Q62219	Q62219 caenorhabdi
19	126.5	4.9	241	11 Q9DDB4	Q9DDB4 mus musculu
20	125	4.8	376	10 Q9STB1	Q9STB1 arabidopsis
21	125	4.8	816	2 Q9Z119	Q9Z119 lactococcus
22	124.5	4.8	317	16 Q51630	Q51630 borrellia bu
23	123.5	4.6	352	5 Q8SUB3	Q8SUB3 encyphallu
24	118.5	4.6	334	17 Q8U0F3	Q8U0F3 pyrococcus
25	117.5	4.6	721	16 Q8YMO5	Q8YMO5 anabaena sp
26	117	4.5	386	10 Q944G2	Q944G2 hevea bras
27	115.5	4.5	4625	10 Q9SMH3	Q9SMH3 chlamydomon
28	115	4.5	816	16 Q9CHS9	Q9CHS9 lactococcus
29	114	4.4	823	5 Q9VMB0	Q9VMB0 diosiphilla
30	114	4.4	3919	16 Q9CPH9	Q9CPH9 pasteurella
31	112.5	4.4	478	16 Q8UHF9	Q8UHF9 agrobacteri
32	112.5	4.4	902	16 Q91742	Q91742 pseudomonas
33	112	4.3	797	17 Q8TPP4	Q8TPP4 methanosaic
34	110.5	4.3	1417	16 Q9HWR8	Q9HWR8 pseudomonas
35	109.5	4.2	1101	4 Q9NXY9	Q9NXY9 homo sapien
36	109.5	4.2	1348	11 Q08828	Q08828 mus musculu
37	109	4.2	3475	5 Q9U175	Q9U175 leishmania
38	108.5	4.2	615	16 Q8UJ10	Q8UJ10 agrobacteri
39	108.5	4.2	659	16 Q8ZHC8	Q8ZHC8 yersinia pe
40	108.5	4.2	1477	2 Q48028	Q48028 haemophilus
41	107	4.1	463	16 Q9S243	Q9S243 streptomyce
42	106.5	4.1	715	2 Q30965	Q30965 rhizobium 1
43	106	4.1	382	10 Q94B09	Q94B09 arabidopsis
44	106	4.1	667	12 Q67482	Q67482 japanese en
45	106	4.1	1087	10 Q82303	Q82303 arabidopsis

# ALIGNMENTS

RESULT 1	ID	Q9C6T1	PRELIMINARY:	PRT:	505 AA.
AC	Q9C6T1				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Hypothetical 54.4 KDa protein.				
GN	F5M6.9.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	Eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=21016719; PubMed=11130712;				
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,				
RA	White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,				
RA	Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,				
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,				
RA	Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,				
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,				
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,				
RA	Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,				
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,				
RA	Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,				
RA	Maltscher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,				
RA	Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,				
RA	Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,				
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,				
RA	Uteckack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,				
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,				
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis				
RL	thaliana.";				
DR	Nature 408:816-820(2000).				
DR	EMBL; AC079041; AAG50716.1;				
DR	TIGRFAMS; TIGR01219; Pnev_kin_ERG8; 1.				

KW Hypothetical protein.  
 SQ SEQUENCE 505 AA; 54409 MW; B1CB6CA338B3D63 CRC64;  
 Query Match 100.0%; Score 2581; DB 10; Length 505;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-190;  
 Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 60  
 DB 1 MAVVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 60  
 QY 61 TSPQLSESMYKLSLNHLTQSVASDSRNPFEVHAIOYIAAAHLATEKDKESLHKL 120  
 DB 61 TSPQLSESMYKLSLNHLTQSVASDSRNPFEVHAIOYIAAAHLATEKDKESLHKL 120  
 QY 121 QGDDITILGNDPFSYRNQIESAGLPLTPESLGTLPAPFASITFNAAESNGANSPEVAKT 180  
 DB 121 QGDDITILGNDPFSYRNQIESAGLPLTPESLGTLPAPFASITFNAAESNGANSPEVAKT 180  
 QY 181 GLGSSAAMTAVVAALLHLYGVVDLSDPCKEKGKFGCSLDVYHMIAGTSHCLAQKVGSG 240  
 DB 181 GLGSSAAMTAVVAALLHLYGVVDLSDPCKEKGKFGCSLDVYHMIAGTSHCLAQKVGSG 240  
 QY 241 FDVSCAVYGSORVYRSPPEVLSPFAQVAVGGLPNEVIGTILGKMDKKRTEFSLPLMN 300  
 DB 241 FDVSCAVYGSORVYRSPPEVLSPFAQVAVGGLPNEVIGTILGKMDKKRTEFSLPLMN 300  
 QY 301 FLGPEGSGSSTPSMVGAVKKMQMDPEKARENQNLSDANLELETKLNDLSKLAKDHW 360  
 DB 301 FLGPEGSGSSTPSMVGAVKKMQMDPEKARENQNLSDANLELETKLNDLSKLAKDHW 360  
 QY 361 VYLKRVKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIIRILROMGEAASVPIE 420  
 DB 361 VYLKRVKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIIRILROMGEAASVPIE 420  
 QY 421 ESOTOLDSTMSAEGVLLAGVPGAGFDPAIFATLGDGSGTKLQWASSHNVALLVREDP 480  
 DB 421 ESOTOLDSTMSAEGVLLAGVPGAGFDPAIFATLGDGSGTKLQWASSHNVALLVREDP 480  
 QY 481 HGVCLESQDPRTCITSGVSSIHLE 505  
 DB 481 HGVCLESQDPRTCITSGVSSIHLE 505

RESULT 2  
 Q944G1 PRELIMINARY; PRT; 503 AA.  
 AC 0944G1  
 DT 01-DEC-2001 (TReMBurel. 19, Created)  
 DT 01-DEC-2001 (TReMBurel. 19, Last sequence update)  
 DT 01-JUN-2002 (TReMBurel. 21, Last annotation update)  
 DE Phosphomevalonate kinase.  
 OS Hevea brasiliensis (Para rubber tree).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid1; Malpighiales; Euphorbiaceae; Hevea.  
 OC NCBI\_TaxID=3981;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. RRM600;  
 RA Hallahan D.L., Kelpner-Hrynko N.M.;  
 RT "genes involved in the biosynthesis of isopentenyl diphosphate in the  
 rubber tree Hevea brasiliensis";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF429385; AAL18926.1;  
 DR TIGRFAMs: TIGR01219; Pmev\_kin\_ERG8; 1.  
 KW kinase.  
 SQ SEQUENCE 503 AA; 54171 MW; BE6P91B80B45FF94 CRC64;

Query Match 74.8%; Score 1930.5; DB 10; Length 503;  
 Best Local Similarity 74.1%; Pred. No. 2.7e-140;  
 Matches 374; Conservative 61; Mismatches 67; Indels 3; Gaps 2;

QY 1 MAVVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 60  
 DB 1 MAVVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 60  
 QY 61 TSPQLSESMYKLSLNHLTQSVASDSRNPFEVHAIOYIAAAHLATEKDKESLHKL 119  
 DB 61 TSPQLSESMYKLSLNHLTQSVASDSRNPFEVHAIOYIAAAHLATEKDKESLHKL 120  
 QY 120 LOGDDITILGNDPFSYRNQIESAGLPLTPESLGTLPAPFASITFNAAESNGANSPEVAK 179  
 DB 120 LOGDDITILGNDPFSYRNQIESAGLPLTPESLGTLPAPFASITFNAAESNGANSPEVAK 180  
 QY 180 TGLSSAAMTAVVAALLHLYGVVDLSDPCKEKGKFGCSLDVYHMIAGTSHCLAQKVGSG 239  
 DB 180 TGLSSAAMTAVVAALLHLYGVVDLSDPCKEKGKFGCSLDVYHMIAGTSHCLAQKVGSG 238  
 QY 240 GFDVSCAVYGSORVYRSPPEVLSPFAQVAVGGLPNEVIGTILGKMDKKRTEFSLPLMN 299  
 DB 239 GFDVSCAVYGSORVYRSPPEVLSPFAQVAVGGLPNEVIGTILGKMDKKRTEFSLPLMN 298  
 QY 300 FLGPEGSGSSTPSMVGAVKKMQMDPEKARENQNLSDANLELETKLNDLSKLAKDHW 359  
 DB 299 FLGPEGSGSSTPSMVGAVKKMQMDPEKARENQNLSDANLELETKLNDLSKLAKDHW 358  
 QY 360 DYLRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIIRILROMGEAASVPIE 419  
 DB 359 DYLRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIIRILROMGEAASVPIE 418  
 QY 420 PESOTOLDSTMSAEGVLLAGVPGAGFDPAIFATLGDGSGTKLQWASSHNVALLVRED 479  
 DB 419 PESOTOLDSTMSAEGVLLAGVPGAGFDPAIFATLGDGSGTKLQWASSHNVALLVRED 478  
 QY 480 PHGVCLESQDPRTCITSGVSSIHLE 504  
 DB 479 PHGVCLESQDPRTCITSGVSSIHLE 503

RESULT 3  
 Q9UT88 PRELIMINARY; PRT; 426 AA.  
 AC 09UT88  
 DT 01-MAY-2000 (TReMBurel. 13, Created)  
 DT 01-MAY-2000 (TReMBurel. 13, Last sequence update)  
 DT 01-JUN-2002 (TReMBurel. 21, Last annotation update)  
 DE Putative phosphomevalonate kinase.  
 GN SPAC343.01C.  
 OS Schizosaccharomyces pombe (Fission Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL109739; CAB52264.1;  
 DR TIGRFAMs: TIGR01219; Pmev\_kin\_ERG8; 1.  
 KW kinase.  
 SQ SEQUENCE 426 AA; 47322 MW; 25635B99E7FAAD64 CRC64;

Query Match 19.1%; Score 493; DB 3; Length 426;  
 Best Local Similarity 30.2%; Pred. No. 1.3e-29;  
 Matches 153; Conservative 83; Mismatches 170; Indels 100; Gaps 21;

QY 1 MAVVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 60  
 DB 1 MKVCSAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIVKPIINEVKEPESAMKMTDVKL 51  
 QY 61 TSPQLSESMYKLSLNHLTQSVASDSRNPFEVHAIOYIAAAHLATEKDKE 113  
 DB 52 KSPQFINAEMLYINIDWVSPTRVHQIYENCLEKNRPVQALFYI-NYFSTGR--- 107



QY 114 SLKLLQGLDITLLGSNDYFYNQIESAGLPLTPESLGLTAD---FASTFENAASNG 170  
 Db 108 --OPLCKMODLQVLTQVNNAY--HQPO-----LKPOQ--TSYPKFNFLNCTLG----- 149  
 QY 171 ANSRPEAKTGIGSSAMTAFAVAALHLYGVNLDSPC--KEKFGCSDLDVYHMAQT 228  
 Db 150 ----GVHKTGIGSSAMTISLIGSL--FLSLRLTDTYDGKSLKIDSTKVYIHNLAOI 202  
 QY 229 SHCIAQKVGSGFVSCAVYSGQRVYRSPFVLSFAQV---AATGLPLNEVIGTILKKG 284  
 Db 203 AHCAQKQKVGSGFVGAATWCSCTYRFPDKLILQOLLVPPDEQIKNINFEELKRIYSKK 262  
 QY 285 WDKRFRFSPLPLMLNLFLEPGSGSGSTPSMVGAVKKQMSDPKAREMNOLSDANLEL 344  
 Db 263 WSDV--VFQPLPATYCLMGDV--AGSGSTPGWVKVQOMQKNEPESK---NCFD----- 311  
 QY 345 ETKLNDLSKLAKHMDVYLR--IKSCSVLTSEKVLHATEPINEAIIKELLEAREMLR 402  
 Db 312 -----DLYSRVLSIKNCFL-----SSSLSLSELOSOFRSIRILQOR 347  
 QY 403 IRLRMQGEASVPIEPESOTOLDSTMSAEGVLLAGVPGAGGDAIFATLIGDSG--T 460  
 Db 348 ITV-----EAKYDIPLKOTNILDNIENQLPVIGVGPAGSGFDAOFLAHHTEITE 400  
 QY 461 KLTOAMSSHVLLALVREDPHCYCLE 486  
 Db 401 NVIKTKMDGVVPMVDSPAPDGLAVE 426

RESULT 4  
 Q973B5 ID 0973B5 PRELIMINARY; PRT; 314 AA.

AC 0973B5;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical protein ST0978.  
 GN ST0978.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM 10545 / 7;  
 RX PubMed=11572479;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-T., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Yamagishi A.,  
 RA Oshina T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL: AP000984; BAB65998.1;  
 DR InterPro: IPR001745; GHMPkinase\_ATP.  
 DR Pfam: PF00288; GHMP\_kinases; 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 314 AA; 35307 MW; 89740213AB16335F CRC64;

Query Match 9.1%; Score 235; DB 17; Length 314;  
 Best Local Similarity 23.2%; Pred. No. 6.3e-10;  
 Matches 113; Conservative 64; Mismatches 129; Indels 182; Gaps 19;  
 Db 3 SAPKILMTGISTYV-----FGISHVIAINKRVR-----CDIK----- 36  
 QY 6 SAPKVLMTGISTYVLEKPNAGLVSTNARPAI--VKPINEVPEKSNAMKWTQVLTSP 63  
 Db 3 SAPKILMTGISTYV-----FGISHVIAINKRVR-----CDIK----- 36  
 QY 64 QLSRESMYKLINHLTQSVASDSRNPFEVHAIOYIAAHLATEKDKESLHKLQGL 123  
 Db 37 -----SSNNF-----IFETTYGTFKDK----- 53

QY 124 DITLLGSNDYFYNQIESAGLPLTPESLGLTADPASTTFENAASNGANSPREYAKTGLG 183  
 Db 54 -----ENLEIESVITVFEKFKGSLPPEHVLFFMDKDFQ--IHGR-----KTGLG 95  
 QY 184 SSAMTAFAVAALHLYGVNLDSPCKRGKFGCSDLDVYHMAQTSRLKAGKXGSGRDV 243  
 Db 96 SSSASTVALTACIYYTL-----FRNLNKDEIYLAQKANIIRQIGISGFDI 142  
 QY 244 SCAYVSGSORVYRFPFVLSFAQVAVTGLPLNEVIGTILKGMKDNKFRFSPL-----MN 299  
 Db 143 ASAYVGSIVYRFR-----YDIEKDVSIPLKIGVIE 174  
 QY 300 LFLGEPGSGSGSTPSMVGAVKKQMSDPKAREMNOLSDANLEETRLANDLSKLAKDW 359  
 Db 175 MLGFGIGSFTVNSVAKFIEK---SNNEEFKKVWKYIDEENI--MAIKLILGK----- 224  
 QY 360 DYLRVIRKSCSVLTSEKVLHATEPINEAIIKELLEAREMLRIRILRMQGEASVPIE 419  
 Db 225 -----IEEAI-----EHVRLARFRFLGLAKKIVGVEIE 252  
 QY 420 PESOTOLDSTMSAEGVLLAGVPGAGGDAIFATLIGDSGTRKLQAMSSHVLLALVRED 479  
 Db 253 NEKIRRLIE--MAENDALIALSPGAG--ESVF--VLGKDLSKVKEKEMVIVYELKED 307  
 QY 480 PHGYCLES 487  
 Db 308 -EGLRIEA 314

RESULT 5  
 Q97UL6 ID 097UL6 PRELIMINARY; PRT; 323 AA.

AC 097UL6;  
 DT 01-OCT-2001 (TREMblrel. 18, Created)  
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE (phospho) mevalonate kinase, putative.  
 GN S502988.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweez M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Eranuso G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Thieriault C., Tolstrup N.,  
 RA Garretts R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 RL EMBL: AE006889; AAK43093.1;  
 DR InterPro: IPR001174; Galkkinase.  
 DR InterPro: IPR001745; GHMPkinase\_ATP.  
 DR Pfam: PF00288; GHMP\_kinases; 1.  
 DR PRINTS: PR00960; LMBPPTOIN.  
 KW kinase; Complete proteome.  
 SQ SEQUENCE 323 AA; 36207 MW; D62E48552F5CBDF8 CRC64;

Query Match 8.9%; Score 230; DB 17; Length 323;  
 Best Local Similarity 22.2%; Pred. No. 1.6e-09;  
 Matches 110; Conservative 72; Mismatches 129; Indels 184; Gaps 21;  
 Db 2 IKVAPGKILMTGISTYV-----FGISHVIAINKRVSCLREIKE----- 42  
 QY 3 VVASAPGKVLMTGISTYVLEKPNAGL--VLSTNARYAIVKPINEVPEKSNAMKWTQVXL 60  
 Db 2 IKVAPGKILMTGISTYV-----FGISHVIAINKRVSCLREIKE----- 42  
 QY 61 TSPQLRESM-YKLINHLTQSVASDSRNPFEVHAIOYIAAHLATEKDKESLHKL 119  
 Db 1 ----- 1

```

Db 43 -----KSLFHTSYGHF-----KNSGNEILNSVL-----DTFRERLSQ-L 77
Qy 120 LOGIDITILGSDNPFYSYRNOIESAGLPLTPEISGLTAPASTFENNAESNGANSKPREVAK 179
Db 78 PQGEYIDLNDKERLI-----DGGK 97
Qy 180 TGLGSSAAMTTAVVAALHLHYGVNDSLPCKEKGCSDDLVIHMIAGTSHCLAQGVGS 239
Db 98 TGLGSSAAMTAVVAALHLHYGVNDSLPCKEKGCSDDLVIHMIAGTSHCLAQGVGS 144
Qy 240 GFDVSCAVYGSORVYRFSPEVLSPFAOVAVTGLPLNEVIGTILKGMKNKREPSLPL- 297
Db 145 GFDVSCAVYGSORVYRFSPEVLSPFAOVAVTGLPLNEVIGTILKGMKNKREPSLPL- 176
Qy 298 --MNLFLGEPGSGSSPTSMGAVKKN--QMSDEPKARENMQNSDANLETKLNDLSKL 354
Db 177 GNDVMDLGFGRKSSST-----VGLVKFVEKSNLDDFKEIMRLIDEENY-MAIKLIKIKL 231
Qy 355 --AKDHVDVLRVYKSCSVLTSEKWLHATEPINEAIIKELLEAREMLRIRILMROMGE 412
Db 232 DEAEVH-----IKGRKYLVNIAERI--VGVKLVSKMEELIKI----- 268
Qy 413 AASVPIEPESOTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDSC-TKLTOAMSSHNLV 472
Db 269 -----AEEEGALVALSPGAGGDSIFA--LGNDLNRVREANSKRGIF 308
Qy 473 ALLVREDPHGVCLGS 487
Db 309 IIDVKED-EGRLLES 322

RESULT 6
Q9FD67 ID Q9FD67 PRELIMINARY; PRT; 368 AA.
AC Q9FD67;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Phosphomevalonate kinase.
GN MYAK2.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
OX NCBI_TaxId=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20353468; PubMed=10894743;
RA Wilding E.I., Brown J.R., Bryant A.P., Chalker A.F., Holmes D.J.,
  Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gyrn M.N.,
  "Identification, evolution and essentiality of the mevalonate pathway
  for isopentenyl diphosphate biosynthesis in gram-positive cocci.",
  RT J. Bacteriol. 182:4319-4327(2000).
RL EMBL: AF290093; AAC02442.1;
DR InterPro: IPR001745; GMPase_ATP.
DR InterPro: IPR001459; Mev_gal_kin.
DR Pfam: PF00288; GMP_kinases.1.
DR PRINTS: PR00959; MEVGALKINASE.
DR TIGRfams: TIGR01220; Pmev_kin_Gr_pos. 1.
KW Kinase.
SQ SEQUENCE 368 AA; 40541 MW; 390993036577146D CRC64;

```

Query Match 6.9%; Score 178.5; DB 2; Length 368;  
 Best local Similarity 22.3%; Pred. No. 1.8e-05;  
 Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;

```

Qy 3 VVASAPKVLMTGTYLLEKPNAGLVSTNARFAIVKPIINEEVKPPSMAMKTDVLTLS 62
Db 2 IEVTTPEKFLIAGETAYAVEGHPAIIIVADQFVTVEETDEGSIQAQSSSLPIWMT- 60
Qy 63 POLSRESMYKLSLNLHLTQSVASDSNPFVEHAIOYAIAAHLATEKDESKHLKLLQG 122
Db 61 ----RRNCELVLDIR-----ENPF-----HYVLAIRHL-TEKVAOENKEL--- 96

```

```

Qy 123 LDITILGSDNPFYSYRNOIESAGLPLTPEISGLTAPASTFENNAESNGANSKPREVAK 182
Db 97 -----SEYHLK-----VTSSEDSNG-----RRYGL 117
Qy 183 GSSAAMTTAVVAAL--LHYGVNDSLPCKEKGCSDDLVIHMIAGTSHCLAQGVGS 240
Db 118 GSSGAVTVGVKALNITYDGL-----ENEIRKLSLAHLAVOGN-GSC 161
Qy 241 EDVSCAVYGSORVYRFSPEVLSPFAOVAVTGLPLNEVIGTILKGMKNKREPSLPLN 299
Db 162 GDIAASYCG--WIAFSTPDHWNOKVT-----TETLTDLADMWPELMFPLKVPQRL 215
Qy 300 LFLGEPGSGSSPTSMGAVKKNQMSDEPKARENMQNSDANLETKLNDLSLADHW 359
Db 216 LLIGWTS-PASTSDLDVRY--HQSKEEQAAVEQFLMSRLCVETMINGFN----- 264
Qy 360 DVYLRVYKSCSVLTSEKWLHATEPINEAIIKELLEAREMLRIRILMROMGEASVPI 419
Db 265 -----TGKIS-VIOKQITKRO-----LAEISLITGVIE 294
Qy 420 PESOTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDSC-TKLTOAMSSHNLV 474
Db 295 TEALKNLCDLAESYTG--AKSSGAGGCGGIYVFRQSGILPLMTAMERKGIITPL 348

RESULT 7
Q93K31 ID Q93K31 PRELIMINARY; PRT; 294 AA.
AC Q93K31;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Phosphomevalonate kinase.
GN PMK.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Lactobacillaceae; Lactobacillus.
OX NCBI_TaxId=1587;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=53/7;
RA Smeds A., Purist T., Palva A.;
  "Identification of a gene cluster for the mevalonate pathway in
  Lactobacillus helveticus.",
  RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AJ279018; CAC51372.1;
DR InterPro: IPR001745; GMPase_ATP.
DR Pfam: PF00288; GMP_kinases.1.
DR TIGRfams: TIGR01220; Pmev_kin_Gr_pos. 1.
KW Kinase.
SQ SEQUENCE 294 AA; 32537 MW; 0795099A1BC97938 CRC64;

```

Query Match 6.1%; Score 158.5; DB 2; Length 294;  
 Best local Similarity 24.8%; Pred. No. 0.00045;  
 Matches 75; Conservative 42; Mismatches 117; Indels 69; Gaps 15;

```

Qy 179 KTGIGSSAAMTTAVVAALHLHYGVNDSLPCKEKGCSDDLVIHMIAGTSHCLAQGVGS 238
Db 49 KYGLGSSAAMTAVVAALHLHYGVNDSLPCKEKGCSDDLVIHMIAGTSHCLAQGVGS 94
Qy 239 GFDVSCAVYGS--ORVYRFSPEVLSPFAOVAVTGLPLNEVIGTILKGMKNKREPSLPL 297
Db 95 SAGDIAASYGVGMVATYQTFKKWLY-ELA-----NKTLSVDVNEAMPGLKELLPLPH 147
Qy 298 --MNLFLGEPGSGSSPTSMGAVKKNQMSDEPKAR-ENMQNSDANLETKLNDLSKLA 355
Db 148 DMKLMIG-----WSQKPASTSRVLDVTANAAALNTEYK----- 181
Qy 356 KDHDVYLRVYKSCSVLTSEKWLHATEPINEAIIKELLEAREMLRIRILMROMGEAAS 415
Db 182 -----NFLASSSEC-VLK-----MIVGFKAKNALIKQIIVN-----RHLLQFAKINQ 225
Qy 416 VPIEPESOTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDSC-TKLTOAMSSHNLV 474

```

Db 226 IAIPIRLTOLIKIAEDRGA--AKTSGAGDGGIVITADTDVDALENEMRRRGLPL 283  
 QY 475 LVR 477  
 Db 284 NFR 286

## RESULT 8

ID Q9KMG3 PRELIMINARY; PRT; 374 AA.  
 AC Q9KMG3;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Phosphomevalonate kinase.  
 OS Streptomyces sp. (strain CL190).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=93372;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL190;  
 RX MEDLINE=20353446; PubMed=10894721;  
 RA Takagi M., Kuzuyama T., Takahashi S., Seto H.;  
 RT "A gene cluster for the mevalonate pathway from Streptomyces sp.  
 strain CL190.";  
 RL J. Bacteriol. 182:4153-4157(2000).  
 DR EMBL; AB037666; BAB07792.1;  
 DR InterPro: IPR00169; Aspartase\_site.  
 DR InterPro: IPR001745; GHMPKase\_ATP.  
 DR Pfam: PF00288; GHMP\_kinases.1.  
 DR PRINTS; PR00959; MEVGAALKINASE.  
 DR TIGRFAMS; TIGR01220; Pmev\_kin\_Gr\_pos.1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOMN\_1.  
 KW kinase.  
 SQ SEQUENCE 374 AA; 39281 MW; A2908B33C694E6D CRC64;

Query Match 6.08; Score 155; DB 2; Length 374;  
 Best Local Similarity 19.78; Pred. No. 0.0012;  
 Matches 97; Conservative 68; Mismatches 112; Indels 156; Gaps 19;

QY 3 VVAAPGKVLMTGGLVLEKPNAGLVLTNARFYAIK-----INEEYKPSW 51  
 Db 8 IVRAIPKGLFVAGERYAVDPGNPAIVAVDNRHISTVSDADADGADYVSSDGLPDAV 67  
 QY 52 AMKMTDKLTSPOLSRSMYKLSLNHLTLQSVASDSRNPVEHAIOYAIANAHLATEKD 111  
 Db 68 GMRHMDRLV-----VRDP--DDQQAARSALAHVY--- 95  
 QY 112 KESLHKLLOGLDITILGSDNPFYRNQIESAGLPITRESIGTLAPFASITFNMAESGA 171  
 Db 96 -----SAIEIVKGRLLGREGKVPRLTLTVSSRLHEDG-- 127  
 QY 172 NSKPEVAKTGLGSSAAMTTAVVAALLHYLV--VDLSDECKEGKFGCSLDIVYHMTAQTSH 230  
 Db 128 -----KRFGLGSSGAVYATVAANAFCGLELSTDERFLAMLATAFELD----- 171  
 QY 231 CLAOGKVGSGFVSCAVYGSQRYVRFSEVLISFAQVAVTGLPLNEVIGTILKGMKNRT 290  
 Db 172 -----PKSGGDLAASWTGG--WIAVQAPDRAF-----VLDLARRGVDRITKAPPGASHV 220  
 QY 291 EFSIPLMLNLF-----GEPGSGSSTPSMGAVKAKKMWSDPEKAREWONLSDALELE 345  
 Db 221 R-RLPAPKGLTLEVGWTEP-----ASTASLVS-----DLHRRTRGSASHQREVE 265  
 QY 346 TKLNDLSKLADHDVYLRVLIKCSVLTSEKMWLHATEPINEAIKELLEREAALRI 405  
 Db 266 T-----TTDCVRSATVALESGLD--DTSILHETRRARQELAR--- 299  
 QY 406 LMRONGEASVPIEPESOTQLDSTMSAEGVLLAGVP--GAGGFDFAIFAITLGDSCGTLT- 463

Db 300 ----LDDEVGIGIEPTKRLTALD---AAEAVGAAKPSGAGDGGIALDLDAEASRDITH 352  
 QY 464 --QAMSSHNVAL 474  
 Db 353 VRQWETAGVLP 365

## RESULT 9

ID Q92F01 PRELIMINARY; PRT; 360 AA.  
 AC Q92F01;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical protein lin0012.  
 GN LIN0012.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nennies G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL596163; CAC95245.1;  
 DR ListList; LIN00012;  
 DR InterPro: IPR001745; GHMPKase\_ATP.  
 DR Pfam: PF00288; GHMP\_kinases.1.  
 DR TIGRFAMS; TIGR01220; Pmev\_kin\_Gr\_pos.1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 360 AA; 40110 MW; 0PFDD7B062842584 CRC64;

Query Match 5.98; Score 151.5; DB 16; Length 360;  
 Best Local Similarity 20.78; Pred. No. 0.0021;  
 Matches 100; Conservative 60; Mismatches 145; Indels 177; Gaps 20;

QY 8 PKVLMGTGGLVLEKPNAGLVLTNARFYAIK-----PINEYKP 48  
 Db 14 PKGLVAGEYVAVESGHAITLAVN-RTITLTLESERNELMIPHYENPVPFGLGELP 72  
 QY 49 ESMANKMTDKLTSPOLSRSMYKLSLNHLTLQSVASDSRNPVEHAIOYAIANAHLAT 108  
 Db 73 DGEHWTFL-----AEAINIAT 88  
 QY 109 EKDKESLKLLOGLDITILGSDNPFYRNQIESAGLPITRESIGTLAPFASITFNMAES 168  
 Db 89 -----TFKSGIGELTPVKM-----VITETLIDQ 112  
 QY 169 NGANSKPEVAKTGLGSSAAMTTAVVAALL--HYLGVVDLSDECKEGKFGCSLDIVYHMTA 226  
 Db 113 SG-----AKYGLGSSAATYAVAINALMTKYPEISMLK-----KLA 150  
 QY 227 QTSKCLAOGKVGSGFVSCAVYGSQRYVRFSEVLISFAQVAVTGLPLNEVIGTILKGMW 286  
 Db 151 ALSHLYVVGSCGDIASCWGMIAVYTFPOEWYKH--RLAYKSID-----WFKKEPMP 203  
 QY 287 NKRTFESLPLMLNLFLEPGSGSSTPSMGAVKAKKMWSDPEKAREWONLSDALELET 346  
 Db 204 MLIETLEPEVPTFESVGTGT--PVSTGLVSGIHAFKQED--SKNYOHFL-----T 251

```

OY 347 KLNDLSKLANDHMDVYLRIKSCSVLTSEKWLHATEPINEAIKELLEAREMLRIRIL 406
DB 252 RNEIMK-----QIIQAFH--TKDELLYSSIKERNRILOEL----- 286
OY 407 MRQGEASVPIEPESOTOLDSTMSAEVLLAG--VPGAGGPAIFAIT--LGSQGTQLQ 464
DB 287 -----GTAGVNIETSLKELAD---SAENMGAGKSSGSGGCGIAFSKTELAELKLVN 339
OY 465 AW 466
DB 340 EW 341

RESULT 10
OY 08YAV1 PRELIMINARY; PRT; 359 AA.
AC 08YAV1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Hypothetical protein lmo0012.
GN LMO0012.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
ON NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusnock C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Ertlin K.-D., Esili H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreitz J., Kuhn M., Kunst F., Kurapat G.,
RA Madunio E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordlek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DB EMBL: AU591973; CAC98227.1;
DR Listlist; LMO0012;
DR InterPro: IPR001745; GHMPkinase_ATP.
DR Pfam: PF00288; GHMP_kinases; 1.
DR TIGRFRAMS: TIGR01220; Pmev_kin_Gr_pos; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 359 AA; 39981 MW; 15156C40BD997A1C CRC64;

Query Match 5.8%; Score 149.5; DB 16; Length 359;
Best Local Similarity 20.5%; Pred. No. 0.003;
Matches 99; Conservative 60; Mismatches 146; Indels 177; Gaps 20;

```

```

DB 150 ALSHLVYQNGSCGDIASCYGMIATTPDQEWKH-RLAYISLE-----WFMKEPWP 202
OY 287 NKRTFELPLMLNLFGEPSGSGSTPSNVAVKMOMSPREKARENOMVLSANILET 346
DB 203 MLIETILEEVPFPFVSQWGT-PIVSTCKLYSQIHAFQED-----SKNYQHFL-----T 250
OY 347 KLNDLSKLANDHMDVYLRIKSCSVLTSEKWLHATEPINEAIKELLEAREMLRIRIL 406
DB 251 RNEIMK-----QIIQAFH--TKDELLYSAIKERNRILOEL----- 285
OY 407 MRQGEASVPIEPESOTOLDSTMSAEVLLAG--VPGAGGPAIFAIT--LGSQGTQLQ 464
DB 286 -----GTAGVNIETSLKELAD---SAENMGAGKSSGSGGCGIAFSKTELAELKLVN 338
OY 465 AW 466
DB 339 EW 340

RESULT 11
OY 09FD62 PRELIMINARY; PRT; 361 AA.
AC 09FD62;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Phosphomevalonate kinase.
GN MVAK2.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
ON NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20353468; PubMed=10894743;
RA Wilding E.I., Brown J.R., Bryant A.P., Chalker A.F., Holmes D.J.,
RA Ingraham K.A., Iordanesco S., So C.Y., Rosenberg M., Gwyn M.N.;
RT "Identification, evolution and essentiality of the mevalonate pathway
for isopentenyl diphosphate biosynthesis in gram-positive cocci.";
RL J. Bacteriol. 182:4319-4327(2000).
DB EMBL: AF290095; AAG02447.1;
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR001745; GHMPkinase_ATP.
DR InterPro: IPR001459; Mev_gal_kin.
DR Pfam: PF00288; GHMP_kinases; 1.
DR PRINTS: PR00959; MEVGALKINASE.
DR TIGRFRAMS: TIGR01220; Pmev_kin_Gr_pos; 1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN.1.
KW kinase.
SQ SEQUENCE 361 AA; 39573 MW; 7802796FC68A0B81 CRC64;

Query Match 5.7%; Score 147.5; DB 2; Length 361;
Best Local Similarity 21.2%; Pred. No. 0.0044;
Matches 102; Conservative 64; Mismatches 171; Indels 145; Gaps 19;

```

Db 164 IAAACYGG--WIAFSTFDHFWLQEOETQHSISELLADWPG-----LSIEPLIAPD 213  
QY 298 MNLELPGSGSGSTPSMGAVKKQMSDPEKARENMQNSDANLELETKLNDLSKLAKD 357  
Db 214 LRLIGTGS-PASTSDLVQV-----HRSRED----- 240  
QY 358 HMDVYLIVKISCVLTSEKWLHATEPINEAI--IKE--LLEAREAMLRIILRMONGEA 413  
Db 241 -----KMAVAQFLKKNSTECVNMINGFKENNNTLLIQMRKROLLHDSAI 288  
QY 414 ASVPIDEPESOTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDG--TKLTQANSSHNVL 472  
Db 289 TGVVIEIPALNKLCINLAEOYEGA--AKSGAGGCGIYVDQSGILPIMSAMEKAET 346  
QY 473 AL 474  
Db 347 PL 348

## RESULT 12

Q9FD72 PRELIMINARY; PRT; 358 AA.  
AC Q9FD72;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE Phosphomevalonate kinase.  
GN MYAK2.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20353468; PubMed=10894743;  
RA Wilding E.I., Brown J.R., Bryant A.P., Chalaker A.F., Holmes D.J.,  
RA Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;  
RT "Identification, evolution and essentiality of the mevalonate pathway  
for isopentenyl diphosphate biosynthesis in gram-positive cocci".  
RL J. Bacteriol. 182:4319-4327(2000).  
DR EMBL: AF290091; AAC02437.1;  
DR InterPro: IPR001745; GHMPkinse\_ATP.  
DR InterPro: IPR001459; Mev\_gal\_kin.  
DR Pfam: PF00288; GHMP\_kinases; 1.  
DR PRINTS: PR00959; MEVGALKINASE.  
DR TIGRFRAMS: TIGR01220; Pmev\_kin\_Gr\_pos; 1.  
KW kinase.  
SQ SEQUENCE 358 AA; 39886 MW; 608EBDADF5774EC2 CRC64;

Query Match 5.5%; Score 142.5; DB 2; Length 358;

Best Local Similarity 19.8%; Pred. No. 0.011;

Matches 94; Conservative 72; Mismatches 167; Indels 141; Gaps 20;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGVLSTNARFYAIVKPIINEEVKPSMAKWTDVKITS 62  
Db 2 IGVKAPGKLYIAGGYAVTEPGYKSILAVNRFVATTEASNK-----VEGSIHS 50  
QY 63 POLSRESM-YKLSLNHLTLOSVSASDSRNPVEHAIOYALIAAHLATEKDESLHKLLQ 121  
Db 51 KTLATEPVKFDNRNEDRLDISDVQAKO-----LKYVTAIEV----- 87  
QY 122 GLDITILGSDNDFYSYRNQIESAGLPLTPESLGTLPAPASITFNAESNGANSKPEVAKTG 181  
Db 88 -----FEDYVYVSCNMNKLHFL-----TDSNLADNSG-----QKYG 119  
QY 182 LGSAAATTAVALALHYLVNDLSDPCKEGKFGCSDDLVIHMTAQTSHCLAQKVGSGF 241  
Db 120 LGSAAALVAVVYKALNEFYG-LLELSNL-----YIKLAVIANMKLQSLSSCG- 165  
QY 242 DVSCAVY-GSORVYRFSPEVLSPFAQVAVTGLPLNEVIGTILKGMKDKRTE-FSLPLLMN 299  
Db 166 DIAVSVYSGMLAVSTFDHDMVK-QQMEETS-----VNDVLEKMPGLHIEPLQAPENME 218

QY 300 LLEGPSSGSSSTPSMGAVKKQMSDPEKARENMQNSDANLELETKLNDLSKLAKDH 359  
Db 219 VLIQWTSPPASS-PHYSEVKRLK-SDP-----SFY 247  
QY 360 DVLIVKISCVLTSEKWLHATEPINEAIKELLEAREAMLRI-RILRMONGEAASVPI 418  
Db 248 GDFLDQSHAC-----VESLIQAFETNNIKGVQKIRIRRIILOSMDNEASVEI 295  
QY 419 EPESOTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDSTKLTQAMSSHNVL 472  
Db 296 EFDLKLKLCVDGEXKHGA--SKTSGAGG-----GDCGITTINKVIRIDRNTI 338

## RESULT 13

Q9FD83 PRELIMINARY; PRT; 358 AA.  
AC Q9FD83;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE Phosphomevalonate kinase.  
GN MYAK2.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20353468; PubMed=10894743;  
RA Wilding E.I., Brown J.R., Bryant A.P., Chalaker A.F., Holmes D.J.,  
RA Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;  
RT "Identification, evolution and essentiality of the mevalonate pathway  
for isopentenyl diphosphate biosynthesis in gram-positive cocci".  
RL J. Bacteriol. 182:4319-4327(2000).  
DR EMBL: AF290087; AAC02426.1;  
DR InterPro: IPR001745; GHMPkinse\_ATP.  
DR InterPro: IPR001459; Mev\_gal\_kin.  
DR Pfam: PF00288; GHMP\_kinases; 1.  
DR PRINTS: PR00959; MEVGALKINASE.  
DR TIGRFRAMS: TIGR01220; Pmev\_kin\_Gr\_pos; 1.  
KW kinase.  
SQ SEQUENCE 358 AA; 40203 MW; 67EF432FA155FEF5 CRC64;

Query Match 5.3%; Score 136.5; DB 2; Length 358;

Best Local Similarity 21.5%; Pred. No. 0.03;

Matches 103; Conservative 60; Mismatches 180; Indels 135; Gaps 23;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGVLSTNARFYAIVKPIINEEVKPSMAKWTDVKITS 62  
Db 2 IGVKAPGKLYIAGGYAVTEPGYKSILALDRF--VTATIEEARQ-----YKGT----- 47  
QY 63 POLSRESMYKLSLNHLTLOSVSASDS---RNPVEHAIQYALIAAHLATEKDESLHKLL 119  
Db 48 -----THSKALHNHPVTSRDEDSIYISPHAKOQLNVTATIEFEO----- 90  
QY 120 LOGDITILGSDNDFYSYRNQIESAGLPLTPESLGTLPAPASITFNAESNGANSKPEVAK 179  
Db 91 -----YAKSCDLAMKHFLITIDS-----NDDDSNG-----HK 117  
QY 180 TGLSSAAMTTAVVAAALHYLVNDLSDPCKEGKFGCSDDLVIHMTAQTSHCLAQKVGSG 239  
Db 118 YGLSSAAVAVSVIKVINEFYDM-----KLSNL-YIKLAVIANMKLQSLSSC 164  
QY 240 GFDVSCAVY-GSORVYRFSPEVLSPFAQVAVTGLPLNEVIGTILKGMKDKRTE-FSLPL 297  
Db 165 G-DIAVSVYSGMLAVSTFDHDMVKH-QIEDT-----TVEVVLKRMNGHLIEPLQAPEN 216  
QY 298 MNLELPGSGSGSTPSMGAVKKQMSDPEKARENMQNSDANLELETKLNDLSKLAKD 357  
Db 217 MEVLIGMTGSPASS-PHYSEVKRLK-SDP-----S 245  
QY 358 HMDVYLIVKISCVLTSEKWLHATEPINEAIKELLEAREAMLRIILRMONGEAASVP 417

Db 246 FYGDFLEDSHRC-----VEK-LIHAFTNNIKGVQKVRON-----RTTIQRMDKEATVD 294  
 Qy 418 IEPESQTLDDSTMSAGVLLAGVPGAGFDALFAITLGD-SGSKLTOAMSSHNVAL 474  
 Db 295 IETEKILYLCIDIAEKYHGA--SKTSAGGSGDCITTIINKVDKEKIYDEWTKGINKPL 350

## RESULT 14

Q9KMF7 PRELIMINARY; PRT; 360 AA.  
 AC Q9KMF7;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Phosphomevalonate kinase.  
 GN PMEVK.  
 OS Kitasatospora griseola (Streptomyces griseolosporus).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Kitasatospora.  
 OX NCBI\_Taxid=2064;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Dairi T.;  
 RT "Cloning of a gene cluster encoding enzymes responsible for the  
 RT mevalonate pathway from a terpenoid antibiotic-producing Streptomyces  
 RT strain."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB037907; BAB07819.1;  
 DR InterPro: IPR001745; GHMPkinase.ATP.  
 DR InterPro: IPR001459; Mey\_gal\_kin.  
 DR Pfam: PF00288; GHMP\_kinases.1.  
 DR PRINTS: PR00959; MEVGALKINASE.  
 DR TIGRfams: TIGR01220; pmev\_kin\_gr\_pos; 1.  
 KW kinase.  
 SQ SEQUENCE 360 AA; 37833 MW; FD194D64355C72DE CRC64;

Query Match 5.3%; Score 136.5; DB 2; Length 360;

Best Local Similarity 19.5%; Pred. No. 0.031;  
 Matches 97; Conservative 67; Mismatches 174; Indels 159; Gaps 17;

Qy 2 AVASAPGKYLMTGTYLVEKPNAGLVLTSTNARFYAIKPK-----INEVVPESWAMK 54  
 Db 6 AVTRRAPGKLFVAGEVAYVEPGRALIVNDRVYTVTSNGARPVYSSDIDAGVYHNP 65  
 Qy 55 WTDVKLTSPQSLRESMYKLSLNHLTQVSASDSRNPVEHAIOYAIAMHATEKDKES 114  
 Db 66 WODGRLT-----GSGTTPHYVAAYE--TVARLLAER----- 94  
 Qy 115 LKHLLOGDITTLGSDNFYSYRNQIESAGLPLTPESLGTLPASTTFMAESNGANSK 174  
 Db 95 -----GNSVPLG--WSISSTLHEDG----- 113  
 Qy 175 PEVAKTGLSSAAMTAVVAALHLVGVNLDSPCKEKGCSDDLVIHMIQTSCHLAQ 234  
 Db 114 ---RKGLGSSGAVYATVSAVAHAGL---ELTADEER-----RTALIAS 153  
 Qy 235 GKV---GSGFDVSCAVYSORVYRSPSEVLSFAQVAVTGLPLNEVIGTTLKGWKNRTE 291  
 Db 154 ARIDPBGSGDIDITWGGVIAVR-APD-----RDAVILDTLRQGVDEALRAPMPSFVR 207  
 Qy 292 FSLPPLMNLFLGEPGSGGSTPSMGAVKKWMSDEPKARENNQNSDANLELETNLNDL 351  
 Db 208 LSPRNLCLEVGWGTGNPVST-----SLLTLD 234  
 Qy 352 SKLAKDMDVYLVIRKSCVLTSEKVLNATEPINEAIIEKELAREAMLRIILKROMG 411  
 Db 235 HRTWMSGSAVRYKATGELV-DAAVIALEDDTEGLLRQVRAHENVR---LDDEVG 290  
 Qy 412 EAASVPIEPESQTLDDSTMSAGVLLAGVPGAGFDALFAITLGDGSK- -LTOAM 466  
 Db 291 LGIFTP-----ELFALCALIARAGAAKRSAGGDC--GIALDLDAEARVDRSPILHRW 341  
 Qy 467 SSHNVALLVREDPHGV 483

Db 342 AAAGVLLPLVSPATEGV 358

## RESULT 15

Q99W20 PRELIMINARY; PRT; 358 AA.  
 AC Q99W20;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Phosphomevalonate kinase.  
 GN WVAK2 OR SAV0592 OR SA0549.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_Taxid=158878, 158879;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.aureus (strain Mu50), and S.aureus (strain N315);  
 RX MEDLINE-21311952; PubMed-11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus."  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL: AP003359; BAB56754.1;  
 DR EMBL: AP003131; BAB41781.1;  
 DR InterPro: IPR001745; GHMPkinase.ATP.  
 DR InterPro: IPR001459; Mey\_gal\_kin.  
 DR Pfam: PF00288; GHMP\_kinases.1.  
 DR PRINTS: PR00959; MEVGALKINASE.  
 DR TIGRfams: TIGR01220; pmev\_kin\_gr\_pos; 1.  
 KW kinase; Complete proteome.  
 SQ SEQUENCE 358 AA; 40217 MW; 19071320FA51AE1 CRC64;

Query Match 5.2%; Score 134.5; DB 16; Length 358;

Best Local Similarity 21.8%; Pred. No. 0.043;  
 Matches 104; Conservative 59; Mismatches 181; Indels 133; Gaps 23;

Qy 3 VVASAPGKYLMTGTYLVEKPNAGLVLTSTNARFYAIKPKINEVVPESWAMKTDVKLS 62  
 Db 2 IQKAPGKLYIAGEVAVTE-PCYKSVLIALDR-----VT 35  
 Qy 63 POLSRESMYKLSLNHLTQVSASDSRNPVEHAIOYAIAMHATEKDKESLHKLLOG 122  
 Db 36 ATTEADQYKGTHTSKALHNHNPVTFSRD--EDSI--VISDPHAAQ----- 77  
 Qy 123 LDTTLGSDNFYSYRNQIESA--GLPLTPESLGTLPASTTFMAESNGANSKPREVAKT 180  
 Db 78 LNVVTAIEIFEOYANSCDIAMKHFLLTDS-----NLDDSG-----HXY 118  
 Qy 181 GLGSSAAMTAVVAALHLVGVNLDSPCKEKGCSDDLVIHMIQTSCHLAQKVGSG 240  
 Db 119 GLGSSAALVSVYKLVNEYTD-----KLSNL-YIYKLVANMKLQSSLSG 165  
 Qy 241 FDVSCAVY-GSORVYRSPSEVLSFAQVAVTGLPLNEVIGTTLKGWKNRTE-FSLPPLM 298  
 Db 166 -DIAVSYSGLMAYSTFDHEWVKH-QIEDT-----TVEEVILKNPGLHIEPLQAPENN 217  
 Qy 299 NLFLEPGSGSGSTPSMGAVKKWMSDEPKARENNQNSDANLELETNLNDLSKLAKH 358  
 Db 218 EVLIGWTGSPASS-PHFVEVVRKLK-SDP-----SF 246  
 Qy 359 MDVYLAVIRKSCVLTSEKVLNATEPINEAIIEKELAREAMLRIILKROMGEASVPI 418  
 Db 247 YGDFLEDSHRC-----VEK-LIHAFTNNIKGVQKVRON-----RTTIQRMDKEATVDI 295

Wed, May 07 14:13:43 2003

us-09-988-863a-2.rsp

Page 9

OY 419 EPESQTOLDSTMSAEVLLAGVPGAGEDAIFATLGD-SCTKLTQAMSHNVLAL 474  
| | | : |||| | | : : :  
Db 296 ETEKLKYLCDIAEKHYNGA--SKTSGAGCGDCGITLINKDVDEKEITYDEMTKHGIKPL 350

Search completed: April 26, 2003, 12:57:02  
Job time : 92 secs

THIS PAGE BLANK (001910)



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:54:00 ; Search time 22 Seconds  
(without alignments)  
2206.722 Million cell updates/sec

Title: US-09-988-863A-2  
Perfect score: 2581  
Sequence: 1 MAVVASAPGKVLMTGTYLV.....ESGDPRTTCITSGVSITLH 505

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR\_73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2581	100.0	505	2 C86443	unknown protein [i
2	493	19.1	426	2 T38650	probable phosphome
3	453.5	17.6	451	2 S57588	phosphomevalonate
4	230	8.9	323	2 P90479	(Phospho) mevalona
5	151.5	5.9	360	2 A61434	mevalonate kinases
6	149.5	5.8	359	2 A61806	mevalonate kinases
7	134.5	5.2	358	2 B89828	phosphomevalonate
8	130	5.0	816	2 T21713	hypothetical prote
9	124.5	4.8	317	2 F70185	phosphomevalonate
10	123	4.8	378	1 S42088	mevalonate kinase
11	117.5	4.6	721	2 A12447	hypothetical prote
12	117	4.5	396	1 A42919	mevalonate kinase
13	117	4.5	557	4 S42226	hypothetical large
14	116	4.5	335	2 A71042	probable mevalonat
15	115	4.5	816	2 B86705	ATP-dependent prot
16	112.5	4.4	478	2 D97447	hypothetical prote
17	112.5	4.4	478	2 A62665	FAD dependent oxid
18	112.5	4.4	902	2 C83635	probable ClpA/B-ty
19	112	4.3	287	1 DXRTBH	libeta-hydroxyster
20	112	4.3	335	2 H75172	libeta-hydroxyster
21	111.5	4.3	395	1 A35629	mevalonate kinase
22	110.5	4.3	1417	1 H83132	probable sensor/re
23	109.5	4.2	292	1 I56604	libeta-hydroxyster
24	109	4.2	662	1 S58298	APPase - pepper (f
25	108.5	4.2	615	2 E98310	rhizobidin secret
26	108.5	4.2	615	2 AF2972	arginine decarboxy
27	108.5	4.2	659	2 AB0114	high-molecular-wel
28	108.5	4.2	1477	2 B43855	probable integral
29	107	4.1	463	2 T36810	

30	107	4.1	1442	2 S57160	sulfite reductase
31	106	4.1	698	1 S31714	MDP-protein ADP-r
32	106	4.1	1087	2 D84767	probable pre-mRNA
33	106	4.1	1436	2 S67655	probable membrane
34	105.5	4.1	1341	2 T17285	hypothetical prote
35	104.5	4.0	397	2 F84545	hypothetical prote
36	104.5	4.0	1983	2 AC1922	two-component hybr
37	104	4.0	574	2 T51799	Clb1-like protein
38	103.5	4.0	284	2 A69536	mevalonate kinase
39	103.5	4.0	291	1 A55573	libeta-hydroxyster
40	103	4.0	710	2 S28014	oud protein - Erw
41	103	4.0	788	2 B84857	hypothetical prote
42	102	4.0	764	2 H82737	malate oxidoreduct
43	102	4.0	812	2 T19446	hypothetical prote
44	101.5	3.9	662	2 T17339	hypothetical prote
45	101	3.9	617	2 AD3123	conserved hypotnet

ALIGNMENTS

RESULT 1

C86443

unknown protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86443

R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hutzler, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Malli, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86443

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <STO>

A:Cross-references: GB:A8005172; NID:G11136726; PIDN:AAG31307.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match	Score	DB 2;	Length	505;
Best Local Similarity	100.0%;	Pred. No. 3.8e-191;		
Matches	505;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MAVVASAPGKVLMTGTYLVLEKPNAGLVSTNAREYATVKNPNEBVKPESNWKMTDVKL	60	
DB	1	MAVVASAPGKVLMTGTYLVLEKPNAGLVSTNAREYATVKNPNEBVKPESNWKMTDVKL	60	
QY	61	TSPQSRSMYKLSLNHLTLOSASDSRNPFEAIOYAIAMAHATEKDESHKLL	120	
DB	61	TSPQSRSMYKLSLNHLTLOSASDSRNPFEAIOYAIAMAHATEKDESHKLL	120	
QY	121	OGDLITILGSNDYFYRNQIESAGLPITPESIGTAPAPASTITFNAESNGANSKREYAKT	180	
DB	121	OGDLITILGSNDYFYRNQIESAGLPITPESIGTAPAPASTITFNAESNGANSKREYAKT	180	
QY	181	GIGSSAAMTTAVVALLHYLGAVDSDPCKEKGPCSDLDVTHMIAQSHCLAQGVSG	240	
DB	181	GIGSSAAMTTAVVALLHYLGAVDSDPCKEKGPCSDLDVTHMIAQSHCLAQGVSG	240	
QY	241	FDVSCAVYGSQRYRVSFVLSFAQVAVTGLPLNEVIGTILGKWDNRRTESLPLNLL	300	
DB	241	FDVSCAVYGSQRYRVSFVLSFAQVAVTGLPLNEVIGTILGKWDNRRTESLPLNLL	300	
QY	301	FLGEGSGSGSSTPSPVAVGAKKQMSDPEKARNQONLSDANILETTKLNDLSKLAKDHD	360	
DB	301	FLGEGSGSGSSTPSPVAVGAKKQMSDPEKARNQONLSDANILETTKLNDLSKLAKDHD	360	

```

Oy      361  YLVAVIVSCSVLTSEKVVNLHATPEINAIIFKELLEAREALRIRIMPOGKAASVPIEP    420
          |||
Db      361  VYLAVIVSCSVLTSEKVVNLHATEPINAIIFKELLEAREALRIRIMROGKAASVPIEP    420
Oy      421  ESQQLDLDTSMAGCVLLAGYPCGAGFDAIFAITLGDSTGLTROAMSNNVALTLVREDP    480
          |||
Db      421  ESQQLDLDTSMAGCVLLAGYPGGGFDFAIFAITLGDSTGLTROAMSSHNVALTLVREDP    480
Oy      481  HGVCLESGDPRTTCITTSGVSSIHLE 505
          |||
Db      481  HGVCLESGDPRTCITTSGVSSIHLE 505

RESULT 2
T38650
probable phosphomevalonate kinase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revise 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: T38650
R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21804
A:Accession: T38650
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-426 <MBR>
A:Cross-references: EMBL:AL109739; PIDN:CAB52264.1; GSPDB:GN00066; SPDB:SPAC343.01c
C:Experimental source: strain 97Zh-, cosmid c343
C:Genetics:
A:Gene: SPDB:SPAC343.01c
A:Map position: 1
A:Introns: 3/2
```

```

Query Match 19.18; Score 493; DB 2; Length 426;
Best Local Similarity 30.28; Pred.No.4.2e-30;
Matches 153; Conservative 83; Mismatches 170; Indels 100; Gaps 21;

QY 1 MAVVASAPGKVLMTGAGLVLEKPNAGLVLTSTNARFYAIRVPIINEEYKPEESMARKWIDVRL 60
D 1 MKVYCSAPAGKVLINGVIVLDPOYSGLVIGLTAKGVASTITLDD-----KCGTAVR 51
QY 61 TSPQ-LSRHSWKYLSL-----HLLQSVASDSRNPVEHAIOYAIAAHLATEKDE 113
D 52 KSPQFINAEMLYINIDVTSPRIHQIYENCLEKPNPFEQLALFYI-NYEFSTGR-- 107
QY 114 SLHKLLLOGLDITILGSDNFYSRNQIESAGCLPLTPESLGTAP---FASITPNAESNG 170
D 108 --QPLCWDDLVLTQVDNAVY-HQPQ-----LKPPQ--TSTPKFNPLCTIG----- 149
QY 171 ANSRKPVAKTGLGSSAAMTTAVVAAALHYLGVDLSDPC--KKGKFGCSOLDYIHMIAOT 228
D 150 -----QVHKRTGLGSSAAMITSLIGSL--FLSLRLTDDTDGDKSLKIDDSIKYIVHMLAQI 202
QY 229 SHCLAQGVSGGFDVSCAVYGSQRYRFRSEVLSFAOV---AVTGLPLNEVIGTILKCK 284
D 203 AHCSSQGVSGGFDVCAATMGSCYRRFRDKLIEQLLVYPDQEIKNINFSTELRKIVSKK 262
QY 285 WDNKRTESLPLRNLTFLGEPGSGGSTPRMGAVKMKQMSDEPKAREMNQNSLDANLEL 344
D 263 WSDV-VPPQLPATYCLLMGDV-AGGSSTPGKVKVQOQMKENPEESK---NEFD----- 311
QY 345 ETKLINDLSKLAKDHWDYLRV--IKSCSVLTSEKWLHAREPINEAIIKELLEAREAMLR 402
D 312 -----DLTSRLYSIKNCFL-----SSESIDSELQSFRIKRIILQR 347
QY 403 IRLMRQMGEAASVPIEPESQTLDDSTMSAEGVLLAGVAGAGFPAIFAITLIGDSG--T 460
D 348 ITV-----EAKVIDEPLKQTNILNDINEQLPVGIVGAGGAGGADQFCIALINHTTEIE 400
QY 461 KLTQAMSSHNVLLALLVREDPHGVCLE 486
D 401 NVIKTKMDGVVPMVDVSPAEGLAVE 426

```

RESULT 3  
S57588  
phosphomevalonate kinase (EC 2.7.4.2) - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YMR959.02; protein YMR220W  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000.  
A:Accession: S57588; A39606  
R:Skellton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S57587.  
A:Accession: S57588  
A:Molecule type: DNA  
A:Residues: 1-451 <SKE>  
A:Cross-references: EMBL:Z49939; NID:9887599; PIDs:9887601; MIPS:YMR220W  
A:Experimental source: Strain AB972  
R:Tsay, Y.H.; Robinson, G.W.  
Mol. Cell. Biol. 11, 620-631, 1991  
A:Title: Cloning and characterization of ERG8, an essential gene of *Saccharomyces cerevisiae*  
A:Reference number: A39606; MUID:91117228; PMID:1846667  
A:Accession: A39606  
A:Molecule type: DNA  
A:Residues: 1-212, 'R', 214-417, 'PLMF', 422, 'D', 424 <TSA>  
A:Cross-references: EMBL:M63648; NID:9553127; PIDs:AAA34596.1; PID:9171479  
C:Genetics:  
A:Gene: SCD:ERG8  
A:Cross-references: SGD:S0004833; MIPS:YMR220W  
A:Map position: 13R  
C:Keywords: phosphotransferase

Query	March Similarity	17.6%	Score 453.5	DB 2	Length 451
Best Local Similarity	30.1%	Pred. No. 5.1e-27			
Matches 153	Conservative	72	Mismatches 165	Indels 103	Gaps 18
QY	6	SAPGCVLMTGGVLYLEKRNAGLIVSTNARFYAIVKPIKE-----EYKPESMAWK---W 55			
Db	8	SAPGKALLAGGLVLDTRYEAFFVGLSARMAHVAHPYSLGSDPKFEVRVSKQKODK 67			
QY	56	TDVKLTSPQLSRESKWKLSLNLHLTQS-----VSASDRKNPFYEHAIOYAIAMAHLATEKD 111			
Db	68	-----LYHISPKSGFIPISIGSGKNPFIEKVIANVFYSYF----- 101			
QY	112	KESLHKLLGGIDITILGSDNPFYSRNQIESAGDLPFRESGLTAPFASITFMAESNGA 171			
Db	102	KPNMDQYCNRLNLEFVIDISDD--AYHSQEDS-----VTEHKG-----NRLSRHS 144			
QY	172	NSKPEVAKTGLGSSAAMTTAAVAALLHYLGVGLDSDPCKEKGFGCSDVDVIHMTAQTSHC 231			
Db	145	HRIEEVPVPTGLGSSAGLVTVLTTLALASF--VSDLENNVDKRR-----EVIHNLAQVAHC 197			
QY	232	LAQGVVSGGFVDSCAVYSSQRTVRFSPVEVLSQAQVAVGLPLNEVYGTITLTKG----- 284			
Db	198	QAQGVIGSGGFVDAALAAAYGISIRYRREPALIS-----NLP---DIGATYSGSKLAHLVD 247			
QY	285	---WNNKRTFESLPLMLNPLIGEPPSGSGSSPMSGAVKAKKMQMSDPKAREKMQWLSAN 341			
Db	248	EDWNNITTKSNHLLPBGCLTLLMKGDI-KNGSEYVKVLQKAKNNYDSIMPESLKYTELDHAN 306			
QY	342	LELEFKLMDLSKLADHDMDVLYLRVIKSC--SVLTSEKVVLLHATEPINEAIIKELLEAREA 399			
Db	307	SRFMGSLKRLDRHLHTHDYDSQDIFESJLERNCCTQK-----PEIETVRDA 353			
QY	400	MURIRILRMQMGESAVPIEPESQTOLLDSYMSAGCVLLAGYPGAGGEDAIFATILGSG 459			
Db	354	VATIRRSFRKITKESGADIEFPVQTSLLDDCQTLTGKGLVITCLLPGAGGDAIAVITKQVD 413			
QY	460	TKLTLQAMSH--NVLALLVREDPHGVCLSGDPRT 492			
Db	414	LNAQTANDKRRPSKVQMLDVTQADMGVRRKK--DPEI 447			

## RESULT 4

F90479  
(Phospho) mevalonate kinase, probable [Imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: F90479  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.  
Submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: F90479  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-323 <KUR>  
A:Cross-references: GB:AE006641; MID:g13816374; PIDN:AAK43093.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SS02988

Query Match 8.9%; Score 230; DB 2; Length 323;  
Best Local Similarity 22.2%; Pred. No. 5, 3e-10;  
Matches 110; Conservative 72; Mismatches 129; Indels 184; Gaps 21;

QY 3 VVASAPKVLMTGGYLVLEKPNAGL--VLSTNAPFAIVKPIINEVKPESWAMKWTQVL 60  
DB 2 IKVASAPKILMIGSYV--FGISHYIAVNRKVSCLREIK----- 42  
QY 61 TSPQLSESM-YKSLNHLITQSVASDSRNPFEHAIQYIAAHLATEKDEKSLHL 119  
DB 43 -----KDSLEFHTSYGF-----KNSGELINSVL-----DFFRRLSO-L 77  
QY 120 LOGIDITILGSNDPFSYRNOIESAGLPLEPSLGLTAPFASITFMAESNGANSKPEVAK 179  
DB 78 PQGEIDLYNDKEII----- 97  
QY 180 TGLGSSAAMTAVVAALLHYLVLDSPCKEGFGCSDDLVIHMAQTSCHLAQKVG 239  
DB 98 TGLGSSAAMTAVVAALLHYLVLDSPCKEGFGCSDDLVIHMAQTSCHLAQKVG 239  
QY 240 GFDVSCAVYGSQRVRFSPPEVLSFAQVAVTGLPLNEVIGITLKKNKPKREFSLPL-- 297  
DB 145 GFDIASAVFSGIVYKRT-----DLDKMPFEKLN 176  
QY 298 --MNLFLGEPGSGSSTPSMVGAVKKN--QMSDPEKARENMONLSDANLETKINDSKL 354  
DB 177 GNYMMLGFTGKSET---VGLVRKFEKSNLDDFKEMILDEBN--MAIKLIKLNKL 231  
QY 355 --AKDHMDVYLRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIILRMQGE 412  
DB 232 DEAVEH-----IKGRKYLNYIAERI--VGKLVSKMEELIKI----- 268  
QY 413 AASVPIPESSOTQLDSTMSAEGVLAGVPGAGFCAIFATLDSGKLTQAMSSHNVL 472  
DB 269 -----AEEGALVNLSPGAGGSDIFA--LGNDLNRVREAWSKRCIF 308  
QY 473 ALVREDPHVGLCES 487  
DB 309 IIDVKED-EGRLRES 322

RESULT 5  
AE134  
mevalonate kinases homolog lln0012 [Imported] - Listeria innocua (strain Clp11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE134  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgeit, O.; Entian, K.D.; Fsihl, H.;  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkhat, G.; Madueno, E.; Maitournam, A.; Ma-  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A.; Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE134  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-360 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CA095245.1; PID:g16412433; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lln0012

Query Match 5.9%; Score 151.5; DB 2; Length 360;  
Best Local Similarity 20.7%; Pred. No. 0.00073;  
Matches 100; Conservative 60; Mismatches 145; Indels 177; Gaps 20;

QY 8 PGKVLMTGGYLVLEKPNAGLVLSTNAPFAIVK-----PINEVKP 48  
DB 14 PGKLVAGEVAVVESGHAILTAVN-RITVLTLEDSENELMIDHYENPSPVIGELKP 72  
QY 49 ESMAMKWTQVLKLTSPQLSESMYKLSLNLITQSVASDSRNPFEHAIQYIAAHLAT 108  
DB 73 DGEHMTF-----AAINAT 88  
QY 109 EKDESLHKLLOGLDITILGSNDPFSYRNOIESAGLPLEPSLGLTAPFASITFMAES 168  
DB 89 -----TFKSEGLETFVKM-----VIFELIDQ 112  
QY 169 NGANSKEPVATGIGSSAAMTAVVAALL--HYLGVDLSPCKEGFGCSDDLVIHMA 226  
DB 113 SG-----AKYGLSSAATAVAVINAMLTFTPEISMLK-----KLA 150  
QY 227 QTSCHLAQKVGSGFDVSCAVYGSQRVRFSPPEVLSFAQVAVTGLPLNEVIGITLKKN 286  
DB 151 ALSHLVYVNGNSCDDIASCMYGMIAVTFPDQEWKHK-RLAKYKSL-----WKNKEWP 203  
QY 287 NKRTFSLPLMNLFLGEPGSGSSTPSMVGAVKKNQMSDPEKARENMONLSDANLETK 346  
DB 204 MJOIETLEPVPFTFSVGTGT-PVSTGKLVQIHAFKQED-----SKNYQHL----- 251  
QY 347 KLNDLSKLAKHMDVYLRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIIL 406  
DB 252 RNEIMK-----QIDAFH--TKDELLSSIKENRIIQL----- 286  
QY 407 MRONGEASVPIPESSOTQLDSTMSAEGVLAG--VPGAGFCAIFAT--LGDGKTLTQ 464  
DB 287 ---GTAGVAVIELSLKELAD---SAENMGAGKSSSGGDDGIFASFKTELAEKLVN 339  
QY 465 AW 466  
DB 340 EW 341

RESULT 6  
AE1806  
mevalonate kinases homolog lmo0012 [Imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE1806  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec-  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgeit, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkhat, G.; Madueno, E.; Maitournam, A.; Ma-  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl-  
A.; Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1806  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-359 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CA098227.1; PID:g164409371; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0012

Query Match 5.8%; Score 149.5; DB 2; Length 359;  
 Best Local Similarity 20.5%; Pred. No. 0.001;  
 Matches 99; Conservative 60; Mismatches 146; Indels 177; Gaps 20;

```

QY 8 PGVLMGCVLVEKRNAGLVLTNRFAIVK-----PINEVPR 48
DB 13 PGKLYAVGEVAVESGHTALTAVN-RYITLTLEDESRNLMIPHYENPWPVGELPK 71
QY 49 ESAMKMTDVKLTSPOLSRSMYKLSLNHLTLQSVASDSRNPFEHAIOYATAAHLAT 108
DB 72 DGEHMTFT-----AEAINIAT 87
QY 109 EKDKESLHKLLLOGLDITLIGSNDYFSYRNOIESAGLPLTPESIGTLAPPASITFNAES 168
DB 88 -----TFLKSEGIETLPVKM-----ILETLDIQ 111
QY 169 NGANSKPEVAKTGLSSAAMTTAVVAAL--HILGVVDLSDPCKEKGFGSDLDVHMIA 226
DB 112 SG-----AKYGLGSSAATVAVINAMTKRFPYISMLK-----KFA-----KLA 149
QY 227 QTHCLAGKVGSGFVSCAVYSGORVVRPSPEVLSFAQVAVTGLPLNEVIGTILGKMD 286
DB 150 ALSHLVYQNGSGCDIASCMYCGMIATTTDQEMVKH-KLAYKSL-----WPKKEPP 202
QY 287 NKRTFESLPLMLFLGEPGSGGSGTPSNVGAVKKQMSDPEKARENOMLSDANLELT 346
DB 203 MLIETLEEVPTFESVGMCT-PVSTGKLVSOIHAFOED-----SKNYOHFL-----T 250
QY 347 KLMDLSKLAKHDVYLRVYKSCSVLTSEKVVHATEPINEAIKELLEAREMLRIL 406
DB 251 RNNELMK-----QIIQAFH--TKDELLYSAIKEKRRILQEL----- 285
QY 407 MRQGEAASVPIPESSOTQLDSTMSAEGVLAG-VPAGGFPALFAIT-LGDSGTRLTQ 464
DB 286 ----GTRKAGNITFSLKELAD---SAENMGAGKSGSGGCGIAFSKTRKELAEKLVN 338
QY 465 AW 466
DB 339 EW 340

```

## RESULT 7

B89828  
 phosphomevalonate kinase [imported] - *Staphylococcus aureus* (strain N315)  
 C:Species: *Staphylococcus aureus*  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: B89828  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: B89828  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-358 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13700484; PIDN:BA841781.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: mvak2

Query Match 5.2%; Score 134.5; DB 2; Length 358;  
 Best Local Similarity 21.8%; Pred. No. 0.015;  
 Matches 104; Conservative 59; Mismatches 181; Indels 133; Gaps 23;

```

QY 3 VVASAPGKVLMTGCVLYLEKPNAGLVLTNRFAIVKPIINEEVKPSAMKWTDVKLT 62
DB 2 IQVAKGKTLTAGEVAVTE-PGYKSVLIALDRF-----VF 35
QY 63 POLSRSMYKLSLNHLTLQSVASDSRNPFEHAIOYATAAHLATEKESLHKLLLOG 122

```

```

DB 36 ATTEADQYKGTIHSKALHNNPTESRD---EDSI--VISDPHAKO----- 77
QY 123 LDITLIGSNDYFSYRNOIESA--GLPLTPESIGTLAPPASITFNAESNGANSKPEVAKT 180
DB 78 LNVYTAIELEFEOYAKSCDIAMKHFLLTDS-----NLDSSG-----HKY 118
QY 181 GLGSSAAMTTAVVAALHLYGVVDLSDPCKEKGFGSDLDVHMIAQTHCLAGKVGSG 240
DB 119 GLGSSAAVYAVSVIKVINEFDM-----KLSNL-YIKKLAVINMKLOJLSGCG 165
QY 241 FDVSCAVY-GSORVVRPSPEVLSFAQVAVTGLPLNEVIGTILGKMDNKRTF-FSLPLM 298
DB 166 -DIAVSVYSGWLAVSTFDHEVMKH-OIEDT-----TVEVLKMPGLHIELOAPENN 217
QY 299 NPLGEPGSGSGSTPSMVGAVKWKQMSDPEKARENOMLSDANLELTSLKSLAKDH 358
DB 218 EVLIGMTGSPASS-PHYSEVTRK-LSDP-----SF 246
QY 359 WDYLVRVYKSCSVLTSEKVVHATEPINEAIKELLEAREMLRILMRQGEAASVPI 418
DB 247 YGDFLEDSSHRC---VEK-LIHAFTNNIKGVQKVRON-----RTIIRMDKEATVDI 295
QY 419 EPESQOTQLDSTMSAEGVLAGVPGAGGPDALFAITLGD-SGTRKLRQANSSHVLA 474
DB 296 ETEKRLYLCDDIAEKHYGA-SKTSAGGDCGCTITINKVDKIKYDEWTKGKIFPL 350

```

## RESULT 8

T21713  
 hypothetical protein F33H2.2 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T21713  
 R:Collage, A.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19463  
 A:Accession: T21713  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-816 <WIL>  
 A:Cross-references: EMBL:Z81526; PIDN:CAB04264.1; GSPDB:GN00019; CESP:F33H2.2  
 A:Experimental source: clone F33H2  
 C:Genetics:  
 A:Gene: CESP.F33H2.2  
 A:Map position: 1  
 A:Intons: 51/1; 101/3; 186/1; 270/3; 328/2; 380/3; 509/2; 751/2  
 C:superfamily: *Caenorhabditis elegans* hypothetical protein F33H2.2

Query Match 5.0%; Score 130; DB 2; Length 816;  
 Best Local Similarity 19.0%; Pred. No. 0.13;  
 Matches 106; Conservative 76; Mismatches 203; Indels 174; Gaps 21;

```

QY 68 ESNYKLSLNHLTLQSVASD-----SRNPFEHAIOYATAAHLATEKDEKESLHKLL 120
DB 63 EQLRSESESLMLXPYHLSIDITVTEMLRSLSPSYA---NLTPMLNTEKSYDLPFTAA 119
QY 121 QGLDITLIGSNDYFSYRNOIESAG-----LPLTPESIGTLAPPASITFNAES 168
DB 120 DAVRLGIGRNOYIELNQCRRSNRKLRSKASARELLPQPTMIVESWMRTWVGILS 179
QY 169 NGANSKPEVAKTGLSSAAMTTAVVAALHLYGVVDLSD-----PCKEKGK----- 214
DB 180 D-VKTLNEEKQOVIDRLDITNVVAIPGLLKYSVTSLYDRGLYFPDVPYDNDYITVA 238
QY 215 -----GCSPLDVHIMIAQTHCLAGKVGSGFVSCAVYSGOR 252
DB 239 DGVNMRVLCIDYFETLLKIFVAIDQTTVLEMSQILHLDLO-LVNALISYFPRLOFARK 297
QY 253 YVR-----FSPEVLSFAQVAVTGLPLNEVIGTILGKMDNKRTF 292
DB 298 RVTGGENLTIHTSWTSNSTVTPMSPISPMSLITSTDELNELFTLLLRG--DDDETD 355
QY 293 SLPLMLNPLGEPGS--GGSSTPSMVGAVKWKQMSDPEKARENOMLSD----- 359

```

Db 356 ALSP-----VSEDSALRASSPNPSLNNSSVESFSLSTQFASDLSSGSGYNNRAAFIFDS 410  
 QY 340 -----ANLELETK-----LNDLSKLAKDHMDVYLRVTKSCSVLT---SEKVVLATE 383  
 Db 411 TLTAFLMGNLSLKGHAVALFEVGLADEQMEKLEQLESYVQFAEGDAQRRSMHATA 470  
 QY 384 PINEAIKELLEAREAL-----RIRILRMQK-----EASVPTEP 420  
 Db 471 LIDS--LKSIRQERADLINGSISITLDEKSRQVLOKSGILVMAAPLSTECATIV-- 526  
 QY 421 ESQTQLDSTMSAEGVLLAGVPGAGFDALFAITLGDG-----GTKLTL----- 463  
 Db 527 -----QSVFLIPPPNAETCSPMFRLSIYSACSSGSPSVLPGLTGLTLTPRV 573  
 QY 464 -QAMSSHNVALLLVREDPH 481  
 Db 574 LQSFQHRHFLVSSTKHEPH 592

## RESULT 9

F70185

phosphomevalonate kinase homolog - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C:Accession: F70185

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A&gt;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MID:98065943; PMID:9403685

A:Accession: F70185

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-317 &lt;RLE&gt;

A:Cross-references: GB:AE001169; GB:AE000783; MID:92688611; PIDN:AAC67030.1; PID:92688611

A:Experimental source: strain B31

## Query Match

Best Local Similarity 4.88; Score 124.5; DB 2; Length 317;

Matches 86; Conservative 65; Mismatches 141; Indels 177; Gaps 19;

QY 6 SAGCKVLMTCGYVLEKPNAGLVSTNARFYAIVKPINEEVKPSMAKMTDVLTSPL 65  
 Db 7 SVGNLLMGEYTLLEKGLALAIKRAFFSK-----KSDSKRF-----F 49  
 QY 66 SRSWMYKLSLNLTLQSVASDSRNPFVEHAIOYAIAAHLATEKDESLKLLQGLDI 125  
 Db 50 SKKK-----KIDFSLIENRSDVFKMFAY----- 74  
 QY 126 TILGSNDIFYRNQIESAGLPPIESIGTLAPFASITFNAAESNGANSKEPVATGIGSS 185  
 Db 75 --LSQNFENLEN-----FAVDVYIDTSNFFFN-----DGTKKGGSS 110  
 QY 186 AAMTTAVVAALL--HYLGVDLSDPCKEKGKSGSDLVHMLAQTSKCL-----AGKV 237  
 Db 111 AVAIGVIGGLFLIHNTNVEKEIFK-----YCELANYSOGGI 151  
 QY 238 GSGEDVSCAVYGSQRYRFSPEVLSFAQVAVTGLPLNEVIGTILKGMKDKRTFESLPL 297  
 Db 152 GSGVDIATSIFFG-----VIEF-----EGGFNPKCRQLGAVEF 184  
 QY 298 MNLFLGPGSGSGSTPBGAVKWKQMSDPEKARENQONSDANLELETKLNDLSKLAKD 357  
 Db 185 NDFL-----MGGIQ--AIKTTTSICEYK----- 207  
 QY 358 HMDVYLRVTKSCSVLTSEKVLHATEPINEAIKELLEAREAMRIRILRMQKEASVVP 417  
 Db 208 HRNSILDFILKCN--LEKKKLVLNNSNS--KSLLSLKRARELGLAIG--EALGVSAALP 262  
 QY 418 IEPE--SQTQLDSTMSAEGVLLAGVPGAGFD--AIFAITLGDGSGTK 461

Db 263 SSFDHLIGQDILIKALGAGNETFLVYRPNIEAFNLKISIVLENIGIK 311

## RESULT 10

S42088

mevalonate kinase (EC 2.7.1.36) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S42088

R:Riou, C.; Jourte, Y.; Lacroste, F.; Karst, F.

submitted to the EMBL Data Library, February 1994

A:Description: Isolation and characterization of a cDNA encoding Arabidopsis thaliana

A:Reference number: S42088

A:Accession: S42088

A:Molecule type: mRNA

A:Residues: 1-378 &lt;RIO&gt;

A:Cross-references: EMBL:X77793; MID:9456613; PIDN:CAA54820.1; PID:9456614

C:Function:

A:Description: catalyzes the reversible phosphorylation of (R)-mevalonic acid by ATP

A:Pathway: isoprenoid biosynthesis

A:Note: this enzyme can also utilize GTP, CTP, and UTP

C:Superfamily: mevalonate kinase

C:Keywords: ATP; isoprenoid biosynthesis; phosphotransferase

F:352/339/Region: ATP binding #status predicted

F:352/Binding site: ATP (Lys) #status predicted

## Query Match

Best Local Similarity 4.88; Score 123; DB 1; Length 378;

Matches 100; Conservative 78; Mismatches 151; Indels 150; Gaps 23;

QY 1 MAVASAPGVMTGCVLEKPNAGLVSTNARFYAIVKPINEEVKPSMAKMTDVL 60  
 Db 1 MEVARAPKILLGHAHVHG-----VAAAIID-----LYTVTL 38  
 QY 61 TSPQLSRESMYKLSLNLTLQSVASDSRN--PEVEHAIOYAIAAHLATEK--DKESLHK 117  
 Db 39 RFLPLPSAENNDRLTLQ--LKDLSLEFSWSLARIKEAIPDSSYLCRSTPASCSEETLK- 94  
 QY 118 LLLQGLDITLIGSNDIFYRNQIESAGLPPIESIGT-LAPF-----ASTFN 164  
 Db 95 -----SIAYL-----VEEQNLPEKMWLSSGISITFLMYLRIIGFNPAIVIN 137  
 QY 165 AASNGANSKPEVAKTGLGSSAAMTTAVVAALLHYLGVDLSDPCKEKGKSCD--LDV 221  
 Db 138 SELPYG-----SELGSSAALCVALLTAALL--ASSISEKTRNGMSLDETUEL 184  
 QY 222 IHMLAQTSKLAQKVGSGFDVSCAVYGSQRYRFSPEVLSFAQVAVTGLPLNEVIGTIL 281  
 Db 185 LNKWAFEGEKIIRGK--PSGIDNTVSAVGN-----MIKFCSEITRLQSNMPLRLMI 234  
 QY 282 KGRKDNKRTESLPLMNLPLGEPGSGGSTPBGAVKWKQMSDPEKARENQONLSAN 341  
 Db 235 --TNTRV-----GRNTKALVSGVSOARVHNPAMSVFAVADIS 272  
 QY 342 LELETKLNDLSKLAKDHMDVYLRVTKSCSVLTSEKVLHATEPINEAIKELLEAREAML 401  
 Db 273 KELAAITQ-----SKDETSV-----ERIKELMEAMQGL 305  
 QY 402 RIRILRMQKEASVVPPIESQQTQLDSTMSAEGVLLAGVPGAGFDALFAITLIGDSGT 460  
 Db 306 -----LSMGVSHSIEAVILTYVHK--LVSKLTGAGGGCCV--LTLLPTGT 348

RESULT 11  
 A12447  
 hypothetical protein al15137 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: A12447  
 R:Kaneo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anaerostipes*  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: A12447  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-721 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BAW6836.1; PID:g17134275; GSPD:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all5137

Query Match 4.6%; Score 117.5; DB 2; Length 721;  
 Best Local Similarity 22.3%; Pred. No. 0.95;  
 Matches 96; Conservative 59; Mismatches 136; Indels 139; Gaps 19;

29 LSNARPY-ATVP-----INEEVKESAMKMTVK----- 59  
 25 LAMNNHYQALAPTPRKALLIGINDYKSSLSGCLTVELOKELLYRFGQATDL 84  
 60 -LTPOLSRSMYKLSLNHLTLOS-----VSASDSRNPEVAHQYAAHAAHATE 109  
 85 TLNEQASREFIEAFLDLTKQAKPDVYVPHSGYGTQLPVESGTLQNAL---VTTD 140  
 110 KDKESLHK-----LLLOGL-DITILGNDFYSYRNOIESAGLPL--TPESLG 153  
 141 ENQADDSQIANTYLEDTLTLRLSLPTDHALVLDTSYFTGTINOPAGIKIRAPESPG 200  
 154 TLAPFASIFPN-----AESNGANSKPEVAKTGLG----- 183  
 201 TRLAAEIDPROQIKNONPEFSPVILSATSDQOQSAKELMGFSAGLTYALTQOLWE 260  
 184 SSAAMTTAV---VAALLHYLGVVDSLDP---CKEGKFGSDLDVIMHIMQTSCHLAQGR 236  
 261 STPATIRVSLSHASISHLQGL---SKQOQGLTKKNGQVATVENLLDRT----- 310  
 237 VSGGFDVSCAVGSGRVRSPVLSFAQVAVGCLPNEVIGITLCKMKMKTEESLPR 296  
 311 -----GAGATILAEEDGKTAQVWLGLP---AOVLENYGNSR--FTLAT 351  
 297 LAMNLFGEPSGSGSTPSMVGAVKQMSDPEKA-----RENNQNS--DANLEL 344  
 352 GEELVLRSLMAGRS--PSRGLLAKQAFSVETITPLQVQGLQENVRVLSNINILAL 409  
 345 ETKLNDLSKL 354  
 410 DTKLERIERV 419

## RESULT 12

mevalonate kinase (EC 2.7.1.36) - human  
 N:Alternate names: MK  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Mar-1997 #text\_change 11-Jun-1999  
 C:Accession: A42919; A58527  
 R:Schaefer, B.L.; Bishop, R.W.; Kratunis, V.J.; Kallinowski, S.S.; Mosley, S.T.; Gibson, R.J. Biol. Chem. 267, 13229-13238, 1992  
 A:Title: Molecular cloning of human mevalonate kinase and identification of a missense mutation  
 A:Reference number: A42919; MUID:92317034; PMID:1377680  
 A:Accession: A42919  
 A:Molecule type: mRNA  
 A:Residues: 1-396 <SCH>  
 A:Cross-references: GB:M88468; NID:9307197; PIDN:AA59362.1; PID:9187561  
 A:Note: sequence extracted from NCBI backbone (NCBI:107743, NCBI:107744)  
 R:Graef, E.; Caselmann, W.H.; Wells, J.; Koshy, R.  
 Oncogene 9, 81-87, 1994  
 A:Title: Insertional activation of mevalonate kinase by hepatitis B virus DNA in a human  
 A:Reference number: S42226; MUID:94134441; PMID:8302606  
 A:Accession: A58527  
 A:Molecule type: mRNA  
 A:Residues: 1-396 <GRA>  
 A:Cross-references: EMBL:X75311; NID:9450345; PIDN:CAA53060.1; PID:9450347

C:Genetics:  
 A:Gene: GDB:MKV  
 A:Cross-references: GDB:134189; OMIM:251170  
 A:Map position: 12pter-12qter  
 A:Note: defects in this gene can cause mevalonic aciduria  
 C:Function:  
 A:Description: catalyzes the reversible phosphorylation of (R)-mevalonic acid by ATP  
 A:Pathway: Isoprenoid biosynthesis  
 A:Note: this enzyme can also utilize GTP, CTP, and UTP  
 C:Superfamily: mevalonate kinase  
 C:Keywords: ATP; isoprenoid biosynthesis; phosphotransferase  
 F:335-357/Region: ATP binding #status predicted  
 F:357/Binding site: ATP (lys) #status predicted

Query Match 4.5%; Score 117; DB 1; Length 396;  
 Best Local Similarity 20.2%; Pred. No. 0.39;  
 Matches 104; Conservative 74; Mismatches 183; Indels 154; Gaps 22;

3 VASAPGKYLMTGTYLYLEKPNAGLYLSTNARYALVKNIEVKEPSNMMKMTDYKLS 62  
 6 LTVSAPGKYLHGEHVVH-GKVALAVSLNLTFLRLQPSN-----GKVDLSL 53  
 63 POLSRSMYKLSLNHLTLOSVSASDSRNPEVAHQYAAHAAHATEKDKESLHKLLLO 122  
 54 PNIGIRAMDVA---RLQSLDLS-----FLEGD-----VTPPSQYEKLR--- 92  
 123 LDITILGNDFYSYRNOIESAGLPL--LTPESIGTLAPASIFMAESNGANSKEVA- 178  
 93 -----EVALGLPDCAVTERLAVLA-FLYLYLSICRQORALPSDIYV 133  
 179 -----KTGIGSSAMTTAVVAALLHYLGVVDSLDPCKES---KKGCDLVYIHIAQT 228  
 134 WSELPPGAGIGSSAAVSVCLAAALLTV--CEELPNLKGDCVNRWTKEDLEINWAO 191  
 229 SHCLAGKVGSGGPDVCAVYVRSPEVLSFAQVAVGCLPNEVIGITLCKMKMKTE 288  
 192 GEMHIGN-PSGVDMNAVSTMG-----ALRYQKGISLSLRPALQILL----- 234  
 289 RTEFSPLPMLNLFGEPSGSGSTPSMVGAVKQMSDPEKARNQNSDANLELEYL 348  
 235 -TNTKTP-----RNTRALVAGVRNRLKFEIYAPLTSIDALSLECEYV 279  
 349 NDISKLAKHMDVLYLVKISCSVLTESEKWLHTEPINEATIKELLEARNAMRILMR 408  
 280 GEMGE-----APABQYLYLELIDMNHILN----- 306  
 409 QMGEAASVPIEPESQTLDTSTMSAGVLLAGVPGAFDAIFATLGDSTKLTQAWSS 468  
 307 -----ALGVGHASLDQLQVLT-RARG-LHSKLTGAGGGGC-GITLLRPGLEQPEVAT 356  
 469 HNVALLVREDPHGVCLESQDPRTTCITSGVSIH 503  
 357 KQALTCGFD-----CLE-----TSIGAPGVSIH 380

## RESULT 13

S42226  
 A:Title: Insertional activation of mevalonate kinase by hepatitis B virus DNA in a hu  
 A:Reference number: S42226; MUID:94134441; PMID:8302606  
 A:Accession: S42226  
 A:Molecule type: mRNA  
 A:Residues: 1-557 <GRA>  
 A:Cross-references: EMBL:X75311; NID:9450345; PIDN:CAA53059.1; PID:9450346  
 C:Keywords: phosphotransferase  
 F:1-157/Region: hepatitis B virus large surface antigen (fragment)  
 F:162-557/Region: human mevalonate kinase (fragment)

Query Match 4.5%; Score 117; DB 4; Length 557;  
 Best Local Similarity 20.2%; Pred. No. 0.68;  
 Matches 104; Conservative 74; Mismatches 183; Indels 154; Gaps 22;

QY 3 VVASPGKVLMTGGIVLEKPNAGVLSTNARFAIVPIEVEKPEPMARKMDVKLTLS 62  
 Db 167 LTVASPGKVLMTGGIVLEKPNAGVLSTNARFAIVPIEVEKPEPMARKMDVKLTLS 62  
 QY 63 POLRESNRYKLSLNLHFLQSVASDSRNPFEHAIQVIAAHLATEKDESKHLILLOG 122  
 Db 215 PNIGIKRMDVA-----RQSLDTS-----FLEQGD-----VTTPTSQVEVKL--- 253  
 QY 123 LDITLLGNDIFYSRNOIESAGLP--LTPESLGLAPFASITFNAESNGANSKPEVA- 178  
 Db 254 -----EVAGLPDDCAVTERLAVLA-FLYLILYSICKRQALPSLDIV 294  
 QY 179 -----KTGLSSAAMTTAAVALLHLIGVVDLSDPCKEG---KFGSDLDVIMIAQT 228  
 Db 295 WSELPPGAGLSSAAYSCVLAALLTV--CEEIPNPLKDGCCVNMWTKEDDELINKMAFQ 352  
 QY 229 SHCLAQKVGSGFVSCAVYGSQRYVRSPFVLSFAQVAVGLPLNEVIGTILKGMKN 288  
 Db 353 GERMIHGN-PSGVDNASTWGG-----ALKRHOCKISLSKRSPLQILL----- 395  
 QY 289 RTEFSLPPLMLFLGEPGSGSSPPSMYGAVKKQMSDPEKARENMQLSDANLELETKL 348  
 Db 396 -TNRKVP-----RMTBALVAGVRNRLKPELIVAPLLTSIDALSICEERYL 440  
 QY 349 NDLKRLKDHMDVLYLRVSKSVLTSEKVLHATEPIHEALIKELLEAREMLRILMR 408  
 Db 441 GEMEE-----AAPPEQVLYLEELIDNQHNLN----- 467  
 QY 409 QMGSAAPLEPESOTQLDSTMSAEVLLAGVPAGGFDAIPATLSDSGKILQAMSS 468  
 Db 468 -----ALGVHASLIDQLCYT-RARG-LHSKILGAGGGGC--GITLLKPLGLEPVEVAT 517  
 QY 469 HNVALLVREDPHGVCESSGDPRTCTITSGVSIH 503  
 Db 518 KQALTSCGFD-----CLE-----TSIGAPGVSIH 541

RESULT 14  
 A71042  
 Probable mevalonate kinase - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000  
 C:Accession: A71042  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekita, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic e  
 A:Reference number: A71000; MUID:98344137; PMID:9679194  
 A:Accession: A71042  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-335 <KMW>  
 A:Cross-references: GB:AP000006; NID:93236133; PIDN:BA30737.1; PID:93258054  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH1625  
 C:Superfamily: galactokinase

Query Match 4.5%; Score 116; DB 2; Length 335;  
 Best Local Similarity 21.9%; Pred. No. 0.36;  
 Matches 102; Conservative 60; Mismatches 124; Indels 180; Gaps 25;

QY 3 VVASPGKVLMTGGIVLEKPNAGVLSTNARFAIVPIEVEKPEPMARKMDVKLTLS 57  
 Db 5 VVASPGKVLMTGGIVLEKPNAGVLSTNARFAIVPIEVEKPEPMARKMDVKLTLS 61  
 QY 58 VKLTPSLRSRSMKLSLNLHFLQSVASDSRNPFEHAIQVIAAHLATEKDESKHL 117

Db 62 VSFSEDKIYFETDYGAKEAVLS-----VYRYAIELAL-----ESDKR----- 99  
 QY 118 LLLOGDITLLGNDIFYSRNOIESAGLPPLPESIGLAPFASITFNAESNGANSKPEV 177  
 Db 100 ---VGIDVST-----TSQIPV-----GA----- 114  
 QY 178 AKTGISAAAMTTAAVALLHLIGVVDLSDPCKEGFCGSDLDVIMIAQTSR---CLAQ 234  
 Db 115 ---GLGSSAAVAATIGAVSRILG-LELS---KE-----ELAKIGHKVELLYQ 155  
 QY 235 GKVGSGFVSCAVYGSQRYV---RSPFVLSFAQVAVGLPLNEVIGTILKGMKNKTE 291  
 Db 156 G-ASSGIDPTVSAVGFLYKQKREP-----LPFME----- 186  
 QY 292 FSLPPLMLFLGEPGSGSSPPSMYGAVKKQMSDPEKARENMQLSDANLELETKLNDL 351  
 Db 187 -LP-----IVGVYGSTG-STKELVAMRKRIEKEPE-----LVPEIILFAMKLVDK 231  
 QY 352 SKLAKDHMDVLYLRVSKSVLTSEKVLHATEPIHEALIKELLEAREMLRILMR-QM 410  
 Db 232 AK-----ELLKLBEEKLTLTGELMNIH 257  
 QY 411 GEASVPIEPESOTQLDSTMSAEVLLAGVPAGGFDAIPATLSDSGKILQAMSS 456  
 Db 258 GULBALGVSTKRLGELYAART-AIGAKLTGAGGGGCMYALAPG 302

RESULT 15  
 B86705  
 ATP-dependent proteinase ATP-binding subunit [Imported] - Lactococcus lactis subsp. 1  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: B86705  
 R:Boletton, A.; Winkler, P.; Manger, S.; Jallion, O.; Malame, K.; Weissenbach, J.; Eh genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: B86705  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-816 <STO>  
 A:Cross-references: GB:AE005176; PID:912723546; PIDN:AAK04740.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: clpC  
 C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 4.5%; Score 115; DB 2; Length 816;  
 Best Local Similarity 21.4%; Pred. No. 1.8;  
 Matches 94; Conservative 68; Mismatches 164; Indels 114; Gaps 19;

QY 92 FEHAIQVIAAHLATEKDESKL---HKL---LLO-----GLDITLLGNDIFYSRNOIE 141  
 Db 82 FSPRAEYMTAASFLAHHNSEAVGTEHLALLOVEDGFGIOLKLIKINIVSLRELE 141  
 QY 142 S-AGPLTPESLIGLAPPA--SIFENAESNG-----ANSKPEVATGSGSAAAMTTA 191  
 Db 142 KRTGKLV-PESKKAVTPSKRMAKGAENSTTPILDSVSSDLTBEARLGLDPMIGREA 200  
 QY 192 VVAALHLYGVVDLSDPCKEGFCGSDLDVIMIAQTSRSHCLAQKVGSGFVSCAVYGSQ 251  
 Db 201 EIDRLIHLSRRTKNNPVLVGPVGKSAITTEGLAQR---IVNGQVPIGL----- 247  
 QY 252 RYVRSPEVLSFAQVAVGLPLNEVIGTILKGMKNKRT---FSLPPLMLFLG--- 304  
 Db 248 ---MNSRIMLNMAV-----VAGTKFRGEFEDRLTAIVEVSSDPVITIPIDELHT 296  
 QY 305 -PGSGGSSTP-----SMVGA-----VKRQMSDPEKARENMQL 337  
 Db 297 IITGAGGDSVNDANILKPLANGDQMGVATITTYHEOKITEDEALERLAIINVDPEP 356  
 QY 338 SD---ANLE-LETKLNDLKLAKDHMDVLYLRVSKSVLTSEKVLHATEPINEA--- 388

Db 357 SPDEAIAILOGLEKFEEDYHQVFTDQAIKSAVMSVRYMTSRKLPDKAIDLDEAAAAY 416  
QY 389 -----IKELLEAR-----EAMLRIRIIMRONGEAA-----SVPIEPES 422  
Db 417 KISVKNQOTKRLDLEKELTFEAOEELSEAVIKLDIKASRTKEKAVEKADKIYKFSYKEDK 476  
QY 423 QTOLLSTMSAEGVLLAGVP 442  
Db 477 ROEVTDOAVVAVASTLTGVP 496

Search completed: April 26, 2003, 12:57:30  
Job time : 26 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:50:50 ; Search time 14 Seconds  
(without alignments)  
1496.110 Million cell updates/sec

Title: US-09-988-863a-2

Perfect score: 2581  
Sequence: 1 MAVVASAPKGVLTMTGYLV.....ESGDPPTCTISGVSSIHLE 505

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	453.5	17.6	451	1	ERG8_YEAST
2	123	4.8	378	1	KIME_ARATH
3	117	4.5	396	1	KIME_HUMAN
4	116	4.5	335	1	KIME_PYRHO
5	113.5	4.4	395	1	KIME_MOUSE
6	112	4.3	287	1	DHIL_RAT
7	112	4.3	335	1	KIME_PYRAB
8	111.5	4.3	395	1	KIME_MOUSE
9	109.5	4.2	292	1	DHIL_MOUSE
10	109	4.2	662	1	FTSH_CAPAN
11	107	4.1	529	1	NXIA_MOUSE
12	107	4.1	1442	1	YJ9F_YEAST
13	106	4.1	698	1	ALT_APT6
14	103.5	4.0	284	1	KIME_ARCFU
15	103	4.0	332	1	LDHA_BRAB
16	103	4.0	504	1	GALL_CANPA
17	103	4.0	710	1	GSOD_ERMCH
18	101	3.9	475	1	PBR_SERMA
19	100	3.9	292	1	DHIL_HUMAN
20	100	3.9	698	1	ALT_BPT2
21	100	3.9	1158	1	RIL4_HUMAN
22	99.5	3.9	443	1	KIME_YEAST
23	99.5	3.9	463	1	YDJN_ECOLI
24	99.5	3.9	1066	1	PK1_YEAST
25	98	3.8	2493	1	CYAA_USTMA
26	97	3.8	292	1	DHIL_SHEEP
27	96.5	3.7	505	1	FIXL_BRAJA
28	96.5	3.7	714	1	FTSH_TOBAC
29	96	3.7	854	1	DISL_HUMAN
30	96	3.7	1163	1	SBCC_CIOAB
31	95	3.7	331	1	LDHA_MACRA
32	95	3.7	893	1	GEFH_HUMAN
33	95	3.7	3432	1	POLG_JAEV1

34	94.5	3.7	400	1	YXXA_BACSU
35	94.5	3.7	735	1	IF2C_GUTH
36	94.5	3.7	2841	1	NFL_MOUSE
37	94	3.6	618	1	MBHL_RHOE
38	94	3.6	1209	1	DNBI_HSVB
39	94	3.6	3432	1	POLG_JAEV5
40	93.5	3.6	515	1	GALL_CANAL
41	93.5	3.6	1157	1	XYNA_THESA
42	93	3.6	712	1	GSOD_ERMCH
43	93	3.6	730	1	EP2_METH
44	93	3.6	763	1	ECHA_PIG
45	92.5	3.6	922	1	DPOL_RICPE

## ALIGNMENTS

RESULT 1  
ERG8\_YEAST STANDARD: PRT: 451 AA.

AC P24521:  
01-MAR-1992 (Rel. 21, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
DE Phosphomevalonate kinase (EC 2.7.4.2).  
GN ERG8 OR YMR220W OR YM9959.02.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE OF 1-424 FROM N.A.  
RX MEDLINE=91117228; PubMed=1846667;  
RA Tsay Y.H., Robinson G.W.;  
RT Cloning and characterization of ERG8, an essential gene of  
RT Saccharomyces cerevisiae that encodes phosphomevalonate kinase.\*;  
RL Mol. Cell. Biol. 11:620-631(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.  
CC -! FUNCTION: ERG8 IS AN ESSENTIAL GENE.  
CC -! CATALYTIC ACTIVITY: ATP + (R)-5-phosphomevalonate = ADP + (R)-5-phosphomevalonate.  
CC -! PATHWAY: SECOND STEP IN ISOPRENTENYL DIPHOSPHATE FORMATION.  
CC -! SUBCELLULAR LOCATION: Cytoplasmic.  
CC -! SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: M63648; AAA34596.1; -  
EMBL: Z49939; CA90191.1; -  
PIR: A39606; A39606.  
SGD: S0004833; ERG8.  
InterPro: IPR001745; GHMPKase\_ATP.  
DR TIGRfam: TIGR01219; Pmev\_kin\_ERG8; 1.  
DR PROSITE: PS00627; GHMP\_KINASES\_ATP; 1.  
KW Transferase; Kinase; ATP-binding; Isoprene biosynthesis;  
KW Sterol biosynthesis.  
NP\_BIND 150 160 ATP (POTENTIAL).  
FT BINDING 183 183 ATP (POTENTIAL).  
FT CONFLICT 213 213 A->R (IN REF. 1).  
FT CONFLICT 418 423 TANDK -> PLMTKD (IN REF. 1).  
SQ SEQUENCE 451 AA; 50455 MW; DB2B6C862153683C CRC64;

Query Match 17.6%; Score 453.5; DB 1; Length 451;  
 Best Local Similarity 30.1%; Pred. No. 1.8e-26;  
 Matches 155; Conservative 72; Mismatches 185; Indels 103; Gaps 18;

6 SAAGKVLMTGGVLEKPNAGLVSTNARFYATKPPINE-----EVKESNAWK---N 55  
 8 SAAGKALLAGVYLVLDKRYEAFVWGLSARMAVAHPGSLGGSDKEVRAKSKQFGKEW 67  
 56 TDVKTLSPOLSRSMYKLSLNHTLOS-----VSASDSRNPFEHAIOYATAAHLATEKD 111  
 68 -----LYHISPKSGFLPVSIGSGSKNFIEKVIANVSYSF----- 101  
 112 KESLHKLLOGDLITLIGSDFTSYRNOISAGLPLTPESLGLAPPAITFNAESNGA 171  
 102 KPMMDYCNMNLFEVDFSD--AYHSQEDS-----VTEHRG-----NRLSFSHS 144  
 172 NSKPEVAKTGLSSAAMTAVVAALLHYLVLDSDPCCKGKGCSDLDVYHMAOTSHC 231  
 145 HRIEVPKTLGSSAGLVLTALASFF-VSDLENNVDYR-----EVIHNAQVAHC 197  
 232 LAQKVGSGFVSCAVYGSQRYVRFSPVELSFAQAVATGLPLNEVIGTILKGR----- 284  
 198 QAQKIGSGFDVAAGSIRYRFPALLS-----NLP--DISATYGSKLHLVD 247  
 285 ---WDKRTFSLPLMLNLFLEPGSGSGSTPSMVGAVKKWQMSDPEKARENNONLSAN 341  
 248 EEDMNTIKSNHLPSCGLTMGDI-KNGSETVKLVQKVMYDSHMPESLKITELDHAN 306  
 342 LELETKINDLSKLAKDHWLYLRKSC--SVLTSEKWLHATFPINEAIKELLEARE 399  
 307 SRMDGSLKDRLETHDDYSDQIFESLERNDCTCKY-----PETTERDA 353  
 400 MLRIILRMQGEAAVYPIEPESQTLDDSTMAEVLGLAGVPGAGFDAIFAITLGDSC 459  
 354 VATIRSFRTITESGADIEPQVOTSLDDCQTLKGLVLTCLIPAGAGVDIAVITKQVD 413  
 460 TKLTQAMSSH--NVALLVREDPHGVCLESQDPT 492  
 414 LRAQTANDKRFKRYOWLDVTOADWGVKKEK-DEPT 447

DB 414 LRAQTANDKRFKRYOWLDVTOADWGVKKEK-DEPT 447

RESULT 2  
 KINE\_ARATH STANDARD; PRT; 378 AA.

AC P46086;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE mevalonate kinase (EC 2.7.1.36) (MK).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID:3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20252529; PubMed=10794536;  
 RA Lluich M.A., Masferrer A., Arro M., Boronat A., Ferrer A.;  
 RT "Molecular cloning and expression analysis of the mevalonate kinase  
 gene from Arabidopsis thaliana.";  
 RL Plant Mol. Biol. 42:365-376(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-  
 phosphate mevalonate.  
 CC -1- ENZYME REGULATION: ITS ACTIVITY IS INHIBITED IN VITRO BY GERANYL  
 PYROPHOSPHATE (GPP) AND FARNESYL PYROPHOSPHATE (FPP) THAT BIND

COMPETITIVELY AT THE ATP-BINDING SITE ON THE ENZYME.  
 -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS, MEVALONATE CATABOLISM.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE  
 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X77793; CAA54820.1;  
 DR EMBL: AF14853; AAD31719.1;  
 DR InterPro: IPR001745; GHMPkinase\_ABP.  
 DR InterPro: IPR001459; meval\_kin.  
 DR Pfam: PF00288; GHMP\_kinases.1.  
 DR PRINTS: PR00959; MEVALKINASE.  
 DR TIGRFAMs: TIGR00549; mevalon\_kin.1.  
 DR PROSITE: PS00627; GHMP\_KINASES\_ATP.1.  
 KW Transferase; Kinase; Sterol biosynthesis; ATP-binding.  
 FT NP\_BIND 141 151  
 SQ SEQUENCE 378 AA; 40643 MW; 87E7E5C78B53C66 CRC64;

Query Match 4.8%; Score 123; DB 1; Length 378;  
 Best Local Similarity 20.9%; Pred. No. 0.078;  
 Matches 100; Conservative 78; Mismatches 151; Indels 150; Gaps 23;

1 MAVVASAPGVMTGVLVEKPNAGLVSTNARFYATKPPINEKPEESNAWKMDVKL 60  
 1 MEKARAPGKIIILGHAHVHG-----ST-----AVAAID-----LYTVTL 38  
 61 TSPQLSRSMYKLSLNHTLOSASDSRN-PVEHAIOYATAAHLATEK--DKESLHK 117  
 39 RFLPSAENNDRLTLO--LKDLSLEFSWLAIRKEAIPDSSILCKSTPASCSEETLK- 94  
 118 LLLQGLDITLIGSDFTSYRNOISAGLPLTPESLGL-LAPF-----ASITFN 164  
 95 -----SIATL-----VEQNLPRKEMWLLSGISITFLMLYRIIGFNATVAVIN 137  
 165 AAESENGANSKPEVAKTGLSSAAMTAVVAALLHYLVLDSDPCCKGKGCSD--LDV 221  
 138 SELPYG-----SGLGSSAALCVALTAL--ASSISEKTRNGWSSIDEETVEL 184  
 222 IHMAQTRSHLAGKVGSGSDVSCAVYGSQRYVRFSPVELSFAQAVATGLPLNEVIGTIL 281  
 185 LNKWAFEGEKTIHGR-PSGIDNTVSAVGN-----MKFSGEITRLQSNMPLMLI 234  
 282 KGWMDKRTFESLPLMLNLFLEPGSGSGSTPSMVGAVKKWQMSDPEKARENNONLSAN 341  
 235 -----TNRV-----GRNTKALVSGSRAVRHPRAMKSVFNAVDSIS 272  
 342 LELETKINDLSKLAKDHWLYLRKSCSVLTSEKWLHATFPINEAIKELLEAREAML 401  
 273 KELAAIIO-----SKDETAY-----TEK-----EERIKELMEMONGLT 305  
 402 RIRILMQGEAAVYPIEPESQTLDDSTMAEVLGLAGVPGAGFDAIFAITLGDSDGT 460  
 306 -----LSMGVSHSIEAVILTYVKKH--LVSKLTGAGGGCV--LTLPLTGT 348

RESULT 3  
 KINE\_HUMAN STANDARD; PRT; 396 AA.

AC Q03426;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE mevalonate kinase (EC 2.7.1.36) (MK).  
 GN MYK.  
 OS Homo sapiens (human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT THR-301.  
 RX MEDLINE=92317034; PubMed=1377680;  
 RA Schaefer B.L., Bishop R.W., Kratunski V.J., Kalinowski S.S.,  
 RA Mosley S.T., Gibson K.M., Tanaka R.D.;  
 RT "Molecular cloning of human mevalonate kinase and identification of a  
 RT missense mutation in the genetic disease mevalonic aciduria.";  
 RL J. Biol. Chem. 267:13229-13238(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatooma;  
 RX MEDLINE=94134441; PubMed=8302606;  
 RA Graef E., Caselmann W.H., Wells J., Koshy R.;  
 RT "Insertional activation of mevalonate kinase by hepatitis B virus DNA  
 RT in a human hepatoma cell line.";  
 RL Oncogene 9:81-87(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A., VARIANTS HIDS P-20; P-39; L-135; T-148; T-268 AND  
 RP I-377, AND VARIANTS MEVALONICACIDURIA P-20; F-264; T-268; T-334 AND  
 RP M-310.  
 RX MEDLINE=21214737; PubMed=11333768;  
 RA Houten S.M., Koster J., Romeijn G.-J., Frenkel J., Di Rocco M.,  
 RA Caruso U., Landrieu P., Kelley R.I., Kuis W., Poll-The B.T.,  
 RA Gibson K.M., Wanders R.J.A., Waterham H.R.;  
 RT "Organization of the mevalonate kinase (MK) gene and identification  
 RT of novel mutations causing mevalonic aciduria and  
 RT hyperimmunoglobulinemia D and periodic fever syndrome.";  
 RL Eur. J. Hum. Genet. 9:253-259(2001).  
 RN [4]  
 RP VARIANTS MEVALONICACIDURIA ILE-243; PHE-264; PRO-265 AND THR-268.  
 RX MEDLINE=99347937; PubMed=10417275;  
 RA Hinson D.D., Ross R.W., Krissans S., Shaw J.L., Kozich V.,  
 RA Rolland M.-O., Divry P., Mancini J., Hoffmann G.F., Gibson K.M.;  
 RT "Identification of a mutation cluster in mevalonate kinase deficiency,  
 RT including a new mutation in a patient of Mennonite ancestry.";  
 RL Am. J. Hum. Genet. 65:327-335(1999).  
 RN [5]  
 RP VARIANTS MEVALONICACIDURIA MET-310 AND THR-334.  
 RX MEDLINE=99330561; PubMed=10401001;  
 RA Houten S.M., Romeijn G.-J., Koster J., Gray R.G.F., Darbyshire P.,  
 RA Smit G.P.A., de Klerk J.B.C., Duran R., Gibson K.M., Wanders R.J.A.,  
 RA Waterham H.R.;  
 RT "Identification and characterization of three novel missense mutations  
 RT in mevalonate kinase cDNA causing mevalonic aciduria, a disorder of  
 RT isoprene biosynthesis.";  
 RL Hum. Mol. Genet. 8:1523-1528(1999).  
 RN [6]  
 RP VARIANTS HIDS PRO-20; THR-268 AND ILE-377.  
 RX MEDLINE=99295935; PubMed=10369261;  
 RA Houten S.M., Kuis W., Duran M., de Koning T.J., van Royen-Kerkhof A.,  
 RA Romeijn G.-J., Frenkel J., Dorland L., de Barse M.M.J.,  
 RA Huijbers W.A.R., Rijkers G.T., Waterham H.R., Wanders R.J.A.,  
 RA Poll-The B.T.;  
 RT "Mutations in MKV, encoding mevalonate kinase, cause  
 RT hyperimmunoglobulinemia D and periodic fever syndrome.";  
 RL Nat. Genet. 22:175-177(1999).  
 RN [7]  
 RP VARIANTS HIDS LEU-167; THR-268 AND ILE-377.  
 RX MEDLINE=99295936; PubMed=10369262;  
 RA Drenth J.P.H., Cuisset L., Grateau G., Vasseur C.,  
 RA van der Velde-Visser S.D., de Jong J.G.N., Beckmann J.S.,  
 RA van der Meer J.W.M., Delpech M.;  
 RT "Mutations in the gene encoding mevalonate kinase cause hyper-IgD and  
 RT periodic fever syndrome.";  
 RL Nat. Genet. 22:178-181(1999).  
 RN [8]  
 RP VARIANTS HIDS N-20; P-39; L-150; L-167; R-202; O-215; T-268;  
 RP S-309; R-326 AND I-377, VARIANT MEVALONICACIDURIA T-334, AND VARIANT  
 RP N-52.  
 RX MEDLINE=21214738; PubMed=1133769;

RA Cuisset L., Drenth J.P.H., Simon A., Vincent M.F.,  
 RA van der Velde-Visser S.D., van der Meer J.W.M., Grateau G.,  
 RA Delpech M.;  
 RT "Molecular analysis of MKV mutations and enzymatic activity in  
 RT hyper-IgD and periodic fever syndrome.";  
 RL Eur. J. Hum. Genet. 9:260-266(2001).  
 CC -1- FUNCTION: MAY BE A REGULATORY SITE IN CHOLESTEROL BIOSYNTHETIC  
 CC PATHWAY.  
 CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-  
 CC phosphomevalonate.  
 CC -1- ENZYME REGULATION: FARNESYL- AND GERANYL-PYROPHOSPHATES ARE  
 CC COMPETITIVE INHIBITORS.  
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS, MEVALONATE CATABOLISM.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and peroxisomal.  
 CC -1- DISEASE: DEFECTS IN MKV ARE THE CAUSE OF MEVALONICACIDURIA. IT IS  
 CC AN ACCUMULATION OF MEVALONIC ACID WHICH CAUSE A VARIETY OF  
 CC SYMPTOMS SUCH AS PSYCHOMOTOR RETARDATION, DYSMORPHIC FEATURES,  
 CC CATARACTS, HEPATOSPLENOMEGALY, LYMPHADENOPATHY, ANEMIA, HYPOTONIA,  
 CC MYOPATHY, AND ATAXIA.  
 CC -1- DISEASE: DEFECTS IN MKV ARE THE CAUSE OF HYPERIMMUNOGLOBULINEMIA D  
 CC AND PERIODIC FEVER SYNDROME (HIDS); AN AUTOSOMAL RECESSIVE DISEASE  
 CC CHARACTERIZED BY RECURRENT EPISODES OF UNEXPLAINED HIGH FEVER  
 CC ASSOCIATED WITH SKIN RASH, DIARRHEA, ADENOPATHY (SWOLLEN, TENDER  
 CC LYMPH NODES), AND ATRALGIAS AND/OR ARTHRITIS. CONCENTRATION OF  
 CC IG, AND OFTEN IGA, ARE ABOVE NORMAL.  
 CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/> or  
 CC send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M88468; AAB59362.1; -;  
 DR EMBL: X75311; CAAS3060.1; -;  
 DR EMBL: X75311; CAAS3059.1; ALT\_INT.  
 DR EMBL: AF217535; AAF82407.1; -;  
 DR EMBL: AF217528; AAF82407.1; JOINED.  
 DR EMBL: AF217529; AAF82407.1; JOINED.  
 DR EMBL: AF217530; AAF82407.1; JOINED.  
 DR EMBL: AF217531; AAF82407.1; JOINED.  
 DR EMBL: AF217532; AAF82407.1; JOINED.  
 DR EMBL: AF217533; AAF82407.1; JOINED.  
 DR EMBL: AF217534; AAF82407.1; JOINED.  
 DR PIR: A42919; A42919.  
 DR GeneW: HGNC:7530; MKV.  
 DR MIM: 251170; -;  
 DR MIM: 260920; -;  
 DR InterPro: IPR001745; GMPKase-ATP.  
 DR InterPro: IPR001459; Mkv\_gal\_kin.  
 DR Pfam: PF002088; GMP\_kinases.1.  
 DR PRINTS: PR00959; MEVALKINASE.  
 DR TIGRFAMs: TIGR00549; mevalon\_kin.1.  
 DR PROSITE: PS00627; GMP\_KINASES-ATP.1.  
 KW Transferase; Kinase; Cholesterol biosynthesis; ATP-binding;  
 KW Peroxisome; Disease mutation; Polymorphism.  
 KW NP\_BIND 138 148  
 FT FT  
 FT VARIANT 20 20  
 FT H -> P (IN HIDS AND MEVALONICACIDURIA).  
 FT /FTID=VAR\_004022.  
 FT VARIANT 39 39  
 FT L -> P (IN HIDS).  
 FT /FTID=VAR\_010957.  
 FT VARIANT 52 52  
 FT S -> N.  
 FT /FTID=VAR\_010958.  
 FT VARIANT 135 135  
 FT S -> L (IN HIDS).  
 FT /FTID=VAR\_010959.  
 FT VARIANT 148 148  
 FT A -> T (IN HIDS).  
 FT /FTID=VAR\_010960.

```

FT VARIANT 150 150 S -> L (IN HIDS).
FT VARIANT 167 167 /FTID=VAR.010961.
FT VARIANT 167 167 P -> L (IN HIDS).
FT VARIANT 202 202 /FTID=VAR.004023.
FT VARIANT 202 202 G -> R (IN HIDS).
FT VARIANT 215 215 /FTID=VAR.010962.
FT VARIANT 215 215 R -> Q (IN HIDS).
FT VARIANT 243 243 /FTID=VAR.010963.
FT VARIANT 243 243 T -> I (IN MEVALONICACIDURIA).
FT VARIANT 264 264 /FTID=VAR.010964.
FT VARIANT 264 264 L -> F (IN MEVALONICACIDURIA).
FT VARIANT 265 265 /FTID=VAR.010965.
FT VARIANT 265 265 L -> P (IN MEVALONICACIDURIA).
FT VARIANT 268 268 /FTID=VAR.010966.
FT VARIANT 268 268 I -> T (IN HIDS AND MEVALONICACIDURIA).
FT VARIANT 301 301 /FTID=VAR.004024.
FT VARIANT 301 301 N -> T (IN MEVALONICACIDURIA; DIMINISHED ACTIVITY).
FT VARIANT 309 309 /FTID=VAR.004025.
FT VARIANT 309 309 G -> S (IN HIDS).
FT VARIANT 310 310 /FTID=VAR.010967.
FT VARIANT 310 310 V -> M (IN MEVALONICACIDURIA).
FT VARIANT 326 326 /FTID=VAR.009068.
FT VARIANT 326 326 G -> R (IN HIDS).
FT VARIANT 334 334 /FTID=VAR.010968.
FT VARIANT 334 334 A -> T (IN MEVALONICACIDURIA).
FT VARIANT 377 377 /FTID=VAR.004026.
FT VARIANT 377 377 V -> I (IN HIDS; MOST FREQUENT MUTATION).
SQ SEQUENCE 396 AA; 42451 MW; C8F6B629B58CD229 CRC64;

```

Query Match 4.5%; Score 117; DB 1; Length 396;  
 Best Local Similarity 20.2%; Pred. No. 0.24;  
 Matches 104; Conservative 74; Mismatches 183; Indels 154; Gaps 22;

```

QY 3 VVASAPKVTMTGTYLVLEKPNAGLVSTNARFYAIKPKINEEKPSPMAKMTDVLTS 62
DB 6 LTVASAPKVTILHGHAAVH-GRVALAVSLNRTFLRLQPSHN-----GKVDLSL 53
QY 63 POLSRESMYKLSLNHLTLOSASDSNRNPFVHAIOYIAAHLATEKDESLHKLLOG 122
DB 54 PNIGIKAMQVA-----RLQSLDLS-----FLEGGD-----VTTPSBOVKLKL 92
QY 123 LDTITLGSNDPYSYRNOIESAGLP--LTPBSLGTLPAPSTTFMAESNGANSKPEVA- 178
DB 93 -----EVAAGLPDDCAVIERLAIVLA-FLLYLSICRQKRALPSLDIV 133
QY 179 -----KTGLSSAAMTTAVVAALHLYGVNDSLPCKEG-----KFGCSLDVYHMTAQT 228
DB 134 WSELTPGAGLSSAAYSVCLAAALTY--CEEJPNPLKDGCVNRKTKEDLELTKWAFO 191
QY 229 SHCLAGKVGSGFVSCAVYGSORRYRFSPEVLSEFAQVAVTGLPLNEVIGTILKGMKDNK 288
DB 192 GERHINHN-PSQVDNNAVSTWG-----ALRKHOGKISLSLRSAIDILL----- 234
QY 289 RTEFSLPPLMNLLEGEFGSSSTPSPVAVKWKQMSDPEKARBNWONLSNDANILETKL 348
DB 235 -TNTKVP-----RNRALVAGVRNRLKTFEPVAVPLTSIDALSICEERYL 279
QY 349 NDLISKLAKHMDVLYRIRKSCSVLTSEKWLHATEPINEMAIKELLEAREMLRIRILMR 408
DB 280 GEMGE-----AAPROYLVLEELIDMNOHHLN----- 306
QY 409 OMGEAASVPIEPESOTOLDSTMSABEVLLAGVPGAGFAIFATITLGDGSKTLQWASS 468
DB 307 -----ALGVGHASLDQLCVT--RARG-LHSKLTGAGGGCC--GITLKGLEQPEVEAT 356
QY 469 HNVALLIVREDPHGVCLSGSDPRTTCITSGVSIH 503
DB 357 KOALTSCGFD-----CLE-----TSIGARGVSIH 380

```

RESULT 4

```

KINE_PYRO ID KINE_PYRO STANDARD; PRT; 335 AA.
AC 059291;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mevalonate kinase (EC 2.7.1.36) (MK).
GN MKV OR PH1625.
OS Pyrococcus horikoshii.
OC Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Onitoku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3." ;
RL DNA Res. 5:55-76(1998).
CC -I- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-
CC -I- phosphomevalonate.
CC -I- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -I- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC SUBFAMILY:
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP000006; BAA30737.1;
CC InterPro: IPR001745; GMPkinse_ATP.
CC InterPro: IPR001459; Mev_gal_kin.
CC Pfam: PF00288; GMP_kinases.1.
CC PRINTS: PR00959; MEVALKINASE.
CC TIGRFS: TIGR00549; mevalon_kin.1.
CC DR PROSITE: PS00627; GMP_KINASES_ATP.1.
CC KW Transferase; Kinase; ATP-binding; Magnesium; Complete proteome.
CC NE_BIND 111 121 ATP (POTENTIAL).
SQ SEQUENCE 335 AA; 35686 MW; 17D0A9E1D22A1EB1 CRC64;

```

Query Match 4.5%; Score 116; DB 1; Length 335;  
 Best Local Similarity 21.9%; Pred. No. 0.22;  
 Matches 102; Conservative 60; Mismatches 124; Indels 180; Gaps 25;

```

QY 3 VVASAPKVTMTGTYLVLEKPNAGLVSTNARFYAIKPKINE-EVKPSPMAKMTD-- 57
DB 5 VVASAPKVTILHGHAAVH-GRVALAVSLNRTFLRLQPSHN-----GKVDLSL 53
QY 58 VKLTPOLSRESMYKLSLNHLTLOSASDSNRNPFVHAIOYIAAHLATEKDESLHKL 117
DB 62 VSESEDKIYETEDYGAKEVLS-----YRYATIELLA-----EESKRA-- 99
QY 118 LLLQGLDITLGSNDPYSYRNOIESAGLP--LTPBSLGTLPAPSTTFMAESNGANSKPEV 177
DB 100 ---VGIDVSI-----TSQIPV-----GA----- 114
QY 178 AKTGIGSSAAMTTAVVAALHLYGVNDSLPCKEGKFGCSLDVYHMTAQTSH---CLAQ 234
DB 115 ---GLSSAANVAVATIGAVSRILG-LELS---KE-----EIALKGHVELLVQ 155
QY 235 GKVGSGFVSCAVYGSORRY--RFSPEVLSEFAQVAVTGLPLNEVIGTILKGMKDNK 291

```

```

Db 156 G-ASSGIDPTVSAVGCFLYKQGFEP-----LPFME----- 186
Qy 292 FSUPLMLNLFGEPSGSSSTPSVYGAVKKQMSDPEKARENNOMLSDANILETKINDL 351
Db 187 --LP-----IYVGYGSGT-STKELVAAYRRKRYEEMPE-----LVEPLLEAMGKLYDK 231
Qy 352 SKLAKDMDVYLRYIKSCSVLTSEKVLHATEPINEAIKIKELLEAREAMLRIRIMR-QM 410
Db 232 AK-----EILSKLDEEKLTKRGLGLMINH 257
Qy 411 GEASVPIEPESQTOLDSTMSAGVLLAGVPGAGFDAIFATLFG 456
Db 258 GLDALGVSTKIKGELVYAARTA-GAIGAKLTGAGGGCWMALAPG 302

RESULT 5
KME_MOUSE
ID KIME_MOUSE STANDARD; PRT; 395 AA.
AC 09R008;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mevalonate kinase (EC 2.7.1.36) (MK).
GN MK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99330561; PubMed=10401001;
RA Houten S.M., Romeijn G.J., Koster J., Gray R.G.F., Darbyshire P.,
RA Smit G.P.A., de Klerk J.B.C., Duran R., Gibson K.M., Wanders R.J.A.,
RA Waterham H.R.,
RT "Identification and characterization of three novel missense mutations
RT in mevalonate kinase cDNA causing mevalonic aciduria, a disorder of
RT isoprene biosynthesis."
RL Hum. Mol. Genet. 8:1523-1528(1999).
CC -|- FUNCTION: MAY BE A REGULATORY SITE IN CHOLESTEROL BIOSYNTHETIC
CC PATHWAY.
CC -|- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-
CC phosphomevalonate.
CC -|- ENZYME REGULATION: FARNESYL- AND GERANYL-PYROPHOSPHATES ARE
CC COMPETITIVE INHIBITORS (BY SIMILARITY).
CC -|- PATHWAY: CHOLESTEROL BIOSYNTHESIS, MEVALONATE CATABOLISM.
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic and peroxisomal.
CC -|- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: AF137598; AF00700.1; -.
DR MGD: MGI:107624; MK.
DR InterPro: IPR001745; GHMPkinse_ATP.
DR InterPro: IPR001174; Galkinase.
DR InterPro: IPR001459; Mev_gal_kin.
DR Pfam: PF00288; GHMP_kinases.1.
DR PRINTS: PR00960; LMBPPROTEIN.
DR PRINTS: PR00959; MEVGALKINASE.
DR TIGRFAMs: TIGR00549; mevalon_kin.1.
DR PROSITE: PS00627; GHMP_KINASES_ATP.1.
DR Transferrase; Kinase; Cholesterol biosynthesis; ATP-binding;
KW Peroxisome.
FT NP_BIND 138 148 ATP (POTENTIAL).
SQ SEQUENCE 395 AA; 41877 MW; 953DB1C89403A3F8 CRC64;

```

```

Query Match 4.4%; Score 113.5; DB 1; Length 395;
Best Local Similarity 21.1%; Pred. No 0.43;
Matches 96; Conservative 75; Mismatches 152; Indels 131; Gaps 22;

Qy 2 AVVASAGKVLMTGGLVLEKPNAGLVLTNARFYATKVPINEEYKESPMANKWTDYKLT 61
Db 5 ALLVASAGKVLHGEHVAHVH-GKVALAALNLPFLILPQSN-----GKYSVN 52
Qy 62 SPQISRESMYKLSLNHLTQSVASDSRNPFVEAIOYATAAHLATEKKEKESLHKLLIO 121
Db 53 LPNIGIKQVMDVGM-----LQRLDRTS-----FLE---OGDVSVPIL-----EQLEKTKM 94
Qy 122 GLDITIGSNDYFYSRNOIESAGLPITPESIGTLAPFASI-----TFNAESNGANSKPE 176
Db 95 G-DLP-----RDRAGNEGMA-----LAFILYLAICRKQRTLPSLDMVYVSELP 139
Qy 177 VAKTGIGSSAAMTTAVVAALLHYVVDLSDPCKEG-----KFGSDLVYHMIQTSHTL 232
Db 140 GA-GLGSSAAYSVCLAALL--TACEEVSNPDKGVSVSRWPEDEKSKTKMAFEGBRV 195
Qy 233 AQGVSGGFYSCAVYGSQRYRRESPEYLSFAQVAVGCLPINEYIGITLKGMKNKRTF 292
Db 196 IHGN-PGVDNAVSTWGA--LRFQOGTMS---SLKSLPSLIQILTNFK----- 238
Qy 293 SLPLMLNLFGEPSGSSSTPSVYGAVKKQMSDPEKARENNOMLSDANILETKINDLS 352
Db 239 -VP-----RSTKALVAAYRSRLTKPELVAPLTSDAISLECEVLEGEM- 282
Qy 353 KLANDMDVYLRYIKSCSVLTSEKVLHATEPINEAIKIKELLEAREAMLRIRIMRQME 412
Db 283 -----VAAPVEQGLVLEELIDMNQHLLN----- 306
Qy 413 AASVPIEPESQTOLDSTMSAGVLLAGVPGAG 446
Db 307 -ALGVGHNSLDLCQYV-AAHG-LHSKLTGAG 336

RESULT 6
DH1L_RAT
ID DH1L_RAT STANDARD; PRT; 287 AA.
AC P16232;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Corticosteroid 11-beta-dehydrogenase, isozyme 1 (EC 1.1.1.146) (11-DH)
DE (11-beta-hydroxysteroid dehydrogenase 1) (11-beta-HSD1).
GN HSD1B1 OR HSD1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=90037015; PubMed=2808402;
RA Agarwal A.K., Monder C., Eckstein B., White P.C.;
RT "Cloning and expression of rat cDNA encoding corticosteroid 11 beta-
RT dehydrogenase."
RL J. Biol. Chem. 264:18939-18943(1989).
RN [2]
RP SEQUENCE OF 1-69 FROM N.A.
RX MEDLINE=92375101; PubMed=1508221;
RA Molisan M.P., Edwards C.R., Seckl J.R.;
RT "Differential promoter usage by the rat 11 beta-hydroxysteroid
RT dehydrogenase gene."
RL Mol. Endocrinol. 6:1082-1087(1992).
RN [3]
RP SEQUENCE OF 1-40.
RX MEDLINE=86278718; PubMed=3460996;
RA Monder C., Shackleton C.H.L., Bradlow H.L., New M.I., Stoner E.,
RA Johan F., Laksini V.;
RT "The syndrome of apparent mineralocorticoid excess: its association

```

RT with 11 beta-dehydrogenase and 5 beta-reductase deficiency and some  
 RT consequences for corticosteroid metabolism.";  
 RL J. Clin. Endocrinol. Metab. 63:550-557(1986).  
 RN [4]  
 RP MUTAGENESIS OF TYR-179 AND LYS-183.  
 RX MEDLINE-93038663; PubMed-1417845;  
 RA Obeid J., White P.C.;  
 RT "Tyr-179 and Lys-183 are essential for enzymatic activity of 11 beta-  
 RT hydroxysteroid dehydrogenase.";  
 RL Biochem. Biophys. Res. Commun. 188:222-227(1992).  
 CC - FUNCTION: CATALYZES REVERSIBLY THE CONVERSION OF CORTICOSTERONE TO  
 CC 11-DEHYDROCORTICOSTERONE.  
 CC - CATALYTIC ACTIVITY: An 11-beta-hydroxysteroid + NADP(+) = an 11-  
 CC oxosteroid + NADPH.  
 CC - SUBCELLULAR LOCATION: Mitochondrial.  
 CC - TISSUE SPECIFICITY: LIVER, KIDNEY, TESTIS AND PLACENTA.  
 CC - PTM: GLYCOSYLATED.  
 CC - SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 CC EMBL: J05107; AAA40886.1; -  
 CC EMBL: S43333; AAB22993.1; -  
 CC PIR: A34430; DXRTH.  
 CC InterPro: IPR002198; ADH\_short.  
 CC Pfam: PF00106; adh\_short.1.  
 CC PROSITE: PS00061; ADH\_SHORT.1.  
 CC Oxidoreductase: Glycoprotein; NADP; Microsome; Steroid metabolism.  
 CC NP\_BIND: 30 78 NADP (BY SIMILARITY).  
 CC ACT\_SITE: 179 179 D->N: SLIGHT LOSS OF ACTIVITY.  
 CC MUTAGEN: 110 110 Y->F/S: COMPLETE LOSS OF ACTIVITY.  
 CC MUTAGEN: 179 179 Y->F/S: COMPLETE LOSS OF ACTIVITY.  
 CC MUTAGEN: 183 183 K->R: COMPLETE LOSS OF ACTIVITY.  
 CC SEQUENCE: 287 AA; 31769 MW; 6E532CFA9ACF1E22 CRC64;  
 SQ  
 Query Match 4.3%; Score 112; DB 1; Length 287;  
 Best Local Similarity 24.6%; Pred. No. 0.35;  
 Matches 52; Conservative 27; Mismatches 82; Indels 50; Gaps 8;  
 QY 43 NEEVRESNAWKTVKLTSPOLSRSMYKLSL--NHLTLOSASDSRNPFEHAIOYA 100  
 DB 20 NEEFREMLOGKKVITGSKGIGREMAHILSKMGAYVL-TARSEGLQKVSRLIEG 78  
 QY 101 IAAAH--LATEKDESLHKL-----LLOGLDITTLG-----SNDFSYRNOIE- 141  
 DB 79 AASAHYIAGTMEADMAFERFVEAGKLGIDMLILNHITQTWMSLFHDIHSRSMVEY 138  
 QY 142 -----SAGCLPLPESLGTLPAPASITFNAAESNGANSKREPAKGTGSSA- 186  
 DB 139 NELSIVLSTAAALPMKOSGSAITAISSMA-----GKMTPLASYSASKFALDGFES 191  
 QY 187 -----AMTTAAVVAALIHLYGVVDLSDPCKE 211  
 DB 192 TIRKEHLMTKVNSITLCVGLFIDTETALKE 222  
 RESULT 7  
 KIME PYRAB  
 ID KIME PYRAB STANDARD; PRT; 335 AA.  
 AC 09V187;  
 DT 16-OCT-2001 (rel. 40, Created)  
 DT 16-OCT-2001 (rel. 40, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)  
 DE Mevalonate kinase (EC 2.7.1.36) (MK).  
 GN MK OR PAB0372.  
 OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_Taxid=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GES / Orsay;  
 RA Heilig R.;  
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome  
 RT structure and evolution.";  
 RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.  
 CC - CATALYTIC ACTIVITY: ATP + (R)-mevalonate -> ADP + (R)-5-  
 CC phosphomevalonate.  
 CC - COFACTOR: MAGNESIUM (BY SIMILARITY).  
 CC - SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC - SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 CC EMBL: AJ248284; CAB49463.1; -  
 CC InterPro: IPR001745; GMPkinase\_ATP.  
 CC InterPro: IPR001174; GMPkinase.  
 CC InterPro: IPR001459; Mv\_gal\_kin.  
 CC Pfam: PF00288; GMP\_Kinases.1.  
 CC PRINTS: PR00960; LMBPPTKIN.  
 CC PRINTS: PR00959; MEVALKINASE.  
 CC TIGRPS: TIGR00549; mevalon\_kin.1.  
 CC PROSITE: PS00627; GMP\_KINASES\_ATP.1.  
 CC Transferase: Kinase; ATP-binding; Magnesium; Complete proteome.  
 CC NP\_BIND: 111 121 ATP (POTENTIAL).  
 CC SEQUENCE: 335 AA; 35774 MW; EDD065DA186599C CRC64;  
 SQ  
 Query Match 4.3%; Score 112; DB 1; Length 335;  
 Best Local Similarity 20.2%; Pred. No. 0.44;  
 Matches 102; Conservative 67; Mismatches 139; Indels 198; Gaps 24;  
 QY 3 VVASAPGKVLMTGCV-LVLEKPNAGLVLTNAPFVIAVKPNE--EVKPESSNAWKTD-- 57  
 DB 5 VLASAPAKIILFEHGSVVYKGP--AIASAIDLFTY-VRAFENSGNIKIEAHDKPGIL 61  
 QY 58 VKLTSPOLSRSMYKLSLNLHITLOSASDSRNPFEHAIOYAIAAHILATEKDESLK 117  
 DB 62 VSESEDKIYFETDYGRAEYLS-----YRHAILEVL-----EADKRT--- 100  
 QY 118 LLLQGLDITILGSDNDFSYRNQIESAGLPPLPESLGTLPAPASITFNAAESNGANSKPREV 177  
 DB 101 -----GVSVSI-----TSQIPV-----GA----- 114  
 QY 178 AKTGGLSSAAMTTAAVVAALIHLYGVVDLSDPCKEKGFGSDLVY--HMTAQTH---CLA 233  
 DB 115 -----GLSSAAVAVAATIGAVSKL-----DLEISKEELAKMGHVELLY 154  
 QY 234 QGKVGSGFDVSCAVVGSQRYRVRSPVLSFAQAAVATGLPLNEVIGTILGKMDKRTERS 293  
 DB 155 QG-ASSGIDPTVSAIGFLY-----KQGEFE 180  
 QY 294 LPTLMNL--FLGPPGSGSPTSMGAVKKMQMSDEKAREMNQNLSDANLELETKLNDL 351  
 DB 181 HLPFVELPIYGVYTGSSG--STKELVAMVRRRYEMRE-----LLEPILSEMGLYVK 231  
 QY 352 SKLAKHMDVYLKRVKSCVLYSEKVVLAHTEPINAIIKELLEAREAMRIILMR-QM 410  
 DB 232 AK-----EVIISK-LDEBEKFLKGLLENNINH 257  
 QY 411 GEASVPIEPESOTOLDSTMSAEGVLLACVPPAGGFDIAFATLTDSDSGKYLQAMSSNH 470

Db 258 GLLDALGVSTKKLSLYAANTA-GAIGAKITGAGGGCMYALAPGK----- 304  
Qy 471 VLALLVREDPHGVCLSESGDPRCTIT 496  
Db 305 -----REVATAIKIAGCTPMITRIS 324

RESULT 8  
KIME RAT  
ID KIME RAT STANDARD: PRT: 395 AA.  
AC p17256:  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mevalonate kinase (EC 2.7.1.36) (MK).  
GN MYK.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=101116;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.  
RC STRAIN=Sorogue-Dawley; TISSUE=Liver;  
RA MEDLINE=90222132; PubMed=2138094;  
RA Tanaka R.D., Lee L.Y., Schafer B.L., Kratunis V.J., Mohler W.A.,  
RA Robinson G.W., Mosley S.T.;  
RT "Molecular cloning of mevalonate kinase and regulation of its mRNA  
RT levels in rat liver."  
RT Proc. Natl. Acad. Sci. U.S.A. 87:2872-2876(1990).  
CC -1- FUNCTION: MAY BE A REGULATORY SITE IN CHOLESTEROL BIOSYNTHETIC  
CC PATHWAY.  
CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-  
CC phosphomevalonate.  
CC -1- ENZYME REGULATION: FARNESYL- AND GERANYL-PYROPHOSPHATES ARE  
CC COMPETITIVE INHIBITORS.  
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS, MEVALONATE CATABOLISM.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and peroxisomal.  
CC -1- DISEASE: MUTATION IN THE MEVALONATE KINASE GENE CAUSES MEVALONIC  
CC ACIDURIA.  
CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
DR EMBL: M29472; AAA41588.1; -  
DR PIR: A35629; A35629.  
DR InterPro: IPR001745; GMPKase-ATP.  
DR InterPro: IPR001459; Mev\_gal\_kin.  
DR Pfam: PF00288; GMP\_kinases; 1.  
DR PRINTS: PRO0059; MEYGALKINASE.  
DR TIGRFAMs: TIGR00549; mevalon\_kin.1.  
DR PROSITE: PS00627; GMP\_KINASES-ATP; 1.  
KW Transferrase; Kinase; Cholesterol biosynthesis; ATP-binding;  
KW Peroxisome.  
FT NP\_BIND 138 148 ATP (POTENTIAL).  
SQ SEQUENCE 395 AA; 41987 MW; 803D1F44E3C525FC CRC64;  
Query Match 4.3%; Score 111.5; DB 1; Length 395;  
Best Local Similarity 20.9%; Pred. No. 0.61;  
Matches 98; Conservative 67; Mismatches 142; Indels 161; Gaps 22.

Qy 3 VVAAPGCVLMTGCVLYLEKPNAGIVLSTNARFAIVAPINEEVKPESMAMKWDVVKLTS 62  
Db 6 LLYVAPGVILHGHAVH -GKVALAVANLRTFLVLRPSN-----GKVSINL 53  
Qy 63 POLSRESMYKLNLNHLTLQSVASDSRNPVEHAIOYAAIAAHLATEKDESLHKLILQG 122

Db 54 PNVGKQVWDVA-----TLQLDPTG-----FLE---QGDVPAPTL-----EQLEKLN 92  
Qy 123 LDITILGNSDFSYNQLSEAGLP---LHPELSTGLA-----PFASITFN 164  
Db 93 -----KVAGLPDPCVGNBGLSLALFLYLALICRQRITLPSLDIMW 134  
Qy 165 AAESNGANSKEVATGTLGSSAAMTTAVVAALLHLYGVVDSLPCKE---GKFGCSDD 220  
Db 135 SELPFGA-----GLGSSAAYSVCVAALL--TACEEVTNPLKDGSGISGWEEDLK 183  
Qy 221 VIHMTAQTSCLIAQGVSGPDVSCAVYGSQRTYRFPSEVLVSFAOAVATGLPLNFYIGTI 280  
Db 184 SINKAAVGERVIHGN-PSGYDNSVTWGA--LRYOOGKMS---SLKRLPALQIILTN 236  
Qy 281 LKGKMDNRTFSPPLNPLNLFLGEGSGSGSSPPSVGVAKKQMSDPKARENMQLSDA 340  
Db 237 TK-----VP-----RSTKALVAGVRSFLIKFPEIMAPLTSIDAI 271  
Qy 341 NLEETKINDSLKAKDHWVLYRIVKSCSVLTSEKWLHATEPINE--AIKELLEARE 398  
Db 272 SLECEKRVGEM-----AAAPVPEQVLYLELMDNQ 302  
Qy 399 AMLRIRILRMQNGEASVPPEPESQTLIDTMSAEGVLLAGVPGAG 446  
Db 303 HHLN-----ALGVGHASLDQLCOYT-AAHG-LHSLKLTGAGG 336

RESULT 9  
DH11.MOUSE  
ID DH11.MOUSE STANDARD: PRT: 292 AA.  
AC P50172;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Corticosteroid 11-beta-dehydrogenase, Isozyme 1 (BC 1.1.1.146) (11-DH)  
DE (11-beta-hydroxysteroid dehydrogenase 1) (11-beta-HSD1) (11beta-  
DE HSD1A).  
GN HSD1A1 OR HSD11.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
RX MEDLINE=95178382; PubMed=7873449;  
RA Rajan V., Chapman K.E., Lyons V., Jamieson P., Mullins J.J.,  
RA Edwards C.R., Seckl J.R.;  
RT "Cloning, sequencing and tissue-distribution of mouse 11 beta-  
RT hydroxysteroid dehydrogenase-1 cDNA."  
RT Eur. J. Biochem. 227:202-208(1995).  
RN [3]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=95154291; PubMed=7851387;  
RA Opperman U.C.T., Netter K.J., Maser E.;  
RT "Cloning and primary structure of murine 11 beta-hydroxysteroid  
RT dehydrogenase/microsomal carbonyl reductase."  
RT Eur. J. Biochem. 227:202-208(1995).  
RN [3]  
RP SEQUENCE OF 1-10 FROM N.A.  
RA Voice M.W.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CATALYZES REVERSIBLY THE CONVERSION OF CORTISOL TO THE  
CC INACTIVE METABOLITE CORTISONE.  
CC -1- CATALYTIC ACTIVITY: An 11-beta-hydroxysteroid + NADP(+) = an 11-  
CC oxosteroid + NADPH.  
CC -1- SUBCELLULAR LOCATION: Microsomal.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN LIVER.  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC (SDR) FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: S75207; AAB3601.1; -  
 DR EMBL: X83202; CA58209.1; -  
 DR EMBL: X92186; CA63096.1; -  
 DR MGD: MG1:103562; Hsd1bl1  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PROSITE: PS00061; ADH\_SHORT; 1.  
 DR Oxidoreductase; NADP; Microsome; Steroid metabolism.  
 FT NP\_BIND 34 63 NADP (BY SIMILARITY).  
 FT ACT\_SITE 183 183 BY SIMILARITY.  
 FT CONFLICT 15 15 F -> S (IN REF. 2).  
 FT CONFLICT 232 232 N -> D (IN REF. 2).  
 FT CONFLICT 234 234 Q -> L (IN REF. 2).  
 FT CONFLICT 261 261 S -> L (IN REF. 2).  
 SQ SEQUENCE 292 AA; 32364 MW; ADE42B1D82DD6CD CRC64;

Query Match 4.28; Score 109.5; DB 1; Length 292;  
 Best Local Similarity 21.7%; Pred. No. 0.55;  
 Matches 70; Conservative 39; Mismatches 109; Indels 105; Gaps 14;

QY 27 LVLTNARFAIVKPIEEVKPEESMAKMTDVKLTSPQSRSMYKSL--NHLTQSVS 84  
 DB 12 LVFLATYYIS-----TNEERPEMLGSKVIYVGASGIGREMAHYHLSKGAHYVL-TAR 66  
 QY 85 ASDSRNPFVEHAIOYAIYAAH--LATEKRESLHKL-----LQGDITILG----- 129  
 DB 67 SEEGIKVRSCLTELGAASAHYIAGTMDMTFAEQFIYKAGKLGMDMLNHIQTSL 126  
 QY 130 ---SNDYSYRNQIE-----SAGLPTEPSLGTLPASFITNAESNANKPE 176  
 DB 127 SLFHDHSHVRVWEVNFSLVYVNSTALPMLKQNSGSIAYISLA-----GKMQPm 179  
 QY 177 VAKTGLSSAA-----MTTAVVAALLHYLGVDLSDCKEKGCCSDLVTHM 224  
 DB 180 IAPYASAKFLDGFSTIRRELYITKYNVSTLCVGLIDETFMKE----- 226  
 QY 225 IAOSSHCLAQKVGSGFDVSCAVYSGQRYVRFPEVLSPFAQVAVTGLPNEVI-GTILKG 283  
 DB 227 ISGIINAQASPK-----EECAL-----ELINGTALR- 252  
 QY 284 KMDNKRTEFSLPLMNLFLGEPG 306  
 DB 253 ---KSEVYYDKSPILTPILGNPG 272

RESULT 10  
 FTSH\_CAPAN STANDARD; PRT; 662 AA.  
 AC Q39444;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cell division protein ftsh homolog, chloroplast precursor  
 DE (EC 3.4.24.-) (Fragment).  
 GN FTSH.  
 OS Capsicum annuum (Bell pepper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; eusterids I; Solanales; Solanaceae; Capsicum.  
 NCBI\_TaxID=4072;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Houline G., Schantz M.L., Schantz R.;  
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.  
 CC -i- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE

CC (BY SIMILARITY).  
 CC -i- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).  
 CC -i- SUBCELLULAR LOCATION: CHLOROPLAST; INTEGRAL MEMBRANE PROTEIN  
 CC (POTENTIAL).  
 CC -i- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
 CC -i- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X90472; CA62084.1; -  
 DR MEROPS: M41.005; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003959; AAA\_ATPase-centr.  
 DR InterPro: IPR003960; AAA\_sub.  
 DR InterPro: IPR000642; Peptidase\_M41.  
 DR Pfam: PF00004; AAA; 1.  
 DR Pfam: PF01434; Peptidase\_M41; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR TIGRfam: TIGR01241; Ftsh\_fam; 1.  
 DR PROSITE: PS00674; AAA; 1.  
 KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;  
 KW Zinc; Chloroplast; Transil peptide.

FT NON\_TER 1 1  
 FT TRANSIT <1 1  
 FT CHAIN 7 662  
 FT TRANSMEM 35 35  
 FT TRANSMEM 173 193  
 FT NP\_BIND 270 277  
 FT METAL 492 492  
 FT ACT\_SITE 493 493  
 FT METAL 496 496  
 FT NON\_TER 662 662  
 SQ SEQUENCE 662 AA; 71061 MW; AF3411BDC502C6A1 CRC64;

Query Match 4.28; Score 109; DB 1; Length 662;  
 Best Local Similarity 20.8%; Pred. No. 2;  
 Matches 119; Conservative 75; Mismatches 209; Indels 168; Gaps 28;

QY 2 AVVASAPGKVLMTGVLLEKPNAG---LVLTNARFYA---IVKPIEEVKPEESMAK 54  
 DB 55 APIASPP-----QVMEVEAPNPNTSNPLPEQNLVLPKQASPSVD--LPESTQWR 105  
 QY 55 WTDVKLTSPQSRSMYKSLNHLTLOSVSASDSRNFVEHAIOYAIYAAH-LATEKRES 114  
 DB 106 YSEF-LNAVKKGKVERVRFSGKGSALO-LTAVDGR-----ANVIYVNDPDL 150  
 QY 115 LHKLLGGLDTTL---GSNDYFYSRNOIESAGLPTEPSLGTLPASFITNAESNGA 171  
 DB 151 IDILAMGVDSISEGSGNGLESVIGNL-----LFPPIADAGLFLFRRSQGG 199  
 QY 172 NSKP-----EVAKTGLGSSAAMTTAVVA---ALLHYLGVD-LSDPC 209  
 DB 200 PGCGGGLGPMDFGRSKSKQVEPETG-----VTFADVAGADQAKLELQVDFLNPD 253  
 QY 210 KEKFGCSDDLVIHMAIOTSHCLAQKVGSGFDV-----SCNV----- 247  
 DB 254 KYTALGAK-----IPKCLLVGPPGKTLTLLARAIVAGEAGVPPFSCAASEVELFV 304  
 QY 248 -YSGQR-----VRFSEVLSTFAQVAVTGL-----PLNEVIGTILKGK 284  
 DB 305 GVGASRVRLHLENKSKAPCIIVFIDEIDAVGRQAGAGGNDEREOTINQLL-TENDGF 363  
 QY 285 WDNK-----RTEFSLPLMNLFLGEGSGGSST-----PSMYGAVKQMSPEKA----- 330  
 DB 364 SGNSGVIVLATATNRPVDLSLLRPGKFDQRYVYDRDVAGRVAILLOVHSGKRLAADVD 423  
 QY 331 -----RENWNLSDANLEETKLNDLSKLAKDH-WDYVLRVI-----KSCSVLT 373



Db 424 FDKIARTPGFTGADLQINMEAAIILAR-RDLKEISKDEISDALERTIACPEKKNAVVS 482  
OY 374 SEKVLAHTEPINEAIKELLEAREAMRILIROMEASVPEPSQOL---LDST 430  
Db 483 DEKKLVAHAGHALVAGALPEYDPAVKISIIIPR--GQAGLTFFAPSERLESGLYSR 540  
OY 431 MSAGVLLAGVAGGAFDAIPA---ITLGS 458  
Db 541 SYLENOMAVALGRVAREVIFGENDVTGAS 571  
RESULT 11  
NXIA\_MOUSE STANDARD: PRT: 529 AA.  
AC 09CS84: 088722;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Neurexin 1-alpha (Neurexin I-alpha) (Fragments).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 1-140 FROM N.A.  
RC STRAIN-CD-1: TISSUE-Brain;  
RA Graveley B.R., Phillips D.L.;  
RT "Sequencing of the neurexin genes."  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 141-529 FROM N.A.  
RC STRAIN-C57BL/6J: TISSUE-Embryo;  
RX MEDLINE-21085660; PubMed-11217851;  
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Pleischmann W., Gaestelund T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojuna N., Carinci P., de Bernaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya K., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlski S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE OF 478-520 FROM N.A.  
RC STRAIN-C57BL/10: TISSUE-Brain;  
RA Gorecki D.C., Sklarczyk A., Lukasik K., Kaczmarek L., Simons J.P.,  
RT "Differential seizure-induced and developmental changes of neurexin  
expression."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NEURONAL CELL SURFACE PROTEIN THAT MAY BE INVOLVED IN  
CELL RECOGNITION AND CELL ADHESION. MAY MEDIATE INTRACELLULAR  
SIGNALING.  
CC -1- SUBUNIT: THE CYTOPLASMIC C-TERMINAL REGION BINDS TO CASK (BY  
SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein (Potential).  
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE ISOFORMS ARE PRODUCED BY  
ALTERNATIVE SPLICING (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS AT LEAST 2 LAMININ G-LIKE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AF387674; AAK70469.1;  
DR EMBL: AF387674; AAK70470.1;  
DR EMBL: AF387674; AAK70471.1;  
DR EMBL: AK017578; BAB30815.1;  
DR EMBL: AJ006802; CAA07257.1;  
DR HSSP: 063373; 1C4R.  
DR MCD: MGI:1096391; Nrxn1.  
DR InterPro: IPR001791; Laminin\_G.  
DR Pfam: PF00054; Laminin\_G; 1.  
DR SMART: SM00282; LaminG; 3.  
DR PROSITE: PS50025; LAM\_G\_DOMAIN; 1.  
KW Transmembrane; Cell adhesion; Repeat; Glycoprotein;  
KW Alternative splicing.  
FT NON\_TER 1  
FT NON\_CONS 140 141  
FT DOMAIN <1 453  
FT TRANSMEM 454 474  
FT DOMAIN 475 529  
FT DOMAIN 15 140  
FT DOMAIN 178 333  
FT CARBOHYD 245 245  
FT VARSPIC 90 96  
FT VARSPIC 82 96  
SQ SEQUENCE 529 AA; 57193 MW; 282E04A03DEBA671 CRC64;  
Query Match 4.18; Score 107; DB 1; Length 529;  
Best Local Similarity 23.48; Pred. No. 2;  
Matches 71; Conservative 41; Mismatches 112; Indels 80; Gaps 16;  
OY 9 GKVLITG---GYLVLEKPNAGLVSTN---AREYIVKPIINEVEYPSMAVKWTFDKLTS 62  
Db 22 GLMLHTGASADYVNLALNKAGVSLVINGSGAFELVPEVCKRFNDNA---WHDVYKT- 76  
OY 63 POLSRSMYKSLMLNTLQSVASDSRNPFEHAIOVAIAAHLATEKDKSLKHLILQG 122  
Db 77 ---NLNQHSGIGHAMVKNLHCS-----VTISVDGLITLTGTQGE----- 113  
OY 123 LDITLGSNDYSTRNQLSAGLPTP--ESIGTLAPR---STFNAESNGANSKPE 176  
Db 114 -DYTMGLGSDDFEYVGSSTADLPSPVCNDPGTYYIFSKGGQITTYWMPDRPSTRAD 172  
OY 177 VAKTGLGSSAMTTPAVNALHLYLGVLDLPCKEGKGCSDLDVYIHMIAGTSHGLAOGK 236  
Db 173 --RLAIGSTYQKEAVLVR-----VDS--GLGDYLELH-----IHGK 208  
OY 237 VGSQFDV-----SCAVYGSORY--VRESPEVLSFAOVAVYGLPINEYIGTLKRW 285  
Db 209 IGVKNVGTDLIAIESNAIINDKXYHVRFRSG--GNATLOVDSP---VIERYPACNN 264  
OY 286 DNKR 289  
Db 265 DNER 268  
RESULT 12  
Y39F\_YEAST STANDARD: PRT: 1442 AA.  
AC P47169;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 161.2 kDa protein in NMD5-HOM6 intergenic region.  
GN YKR137C OR J2126.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;



```

QY 321 KWONS-----DPEKARENMONSDANLEETKINDLSKLAKDHMDVLYRYK 367
DB 426 ECGSGYTDINLLGRNPR-----YDVMSEK--EIESAINLND--AFENGD---RIPE 475
QY 368 SC5VLTSEKWLATEPINEAIKELLEAREAMLRILRONGEASVPI-----418
DB 476 GITYVRAOS--MTADIYALVKNV-----FYRNFVSTLPIIFGRFGITH 521
QY 419 -----EPESOTOL-LDSTMSAEGLV 438
DB 522 AGIGLEPEARNELTVDK--NEECITI 546

RESULT 14
KINE_ARCFU
ID KINE_ARCFU STANDARD; PRT; 284 AA.
AC 027985;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN MEVALONATE KINASE (EC 2.7.1.36) (MK).
SVK OR AF2289.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OX Archaeoglobaceae; Archaeoglobus.
NCBI_TaxID=2234;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Ketch C.J., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate -> ADP + (R)-5-
CC phosphomevalonate.
CC -1- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000946; AAB8965.1;
CC TIGR: AF2289;
CC InterPro: IPR001745; GMPkinase_ATP.
CC InterPro: IPR001459; Mev_gal_kin.
CC Pfam: PF00288; GMP_kinases_1.
CC PRINTS: PR00959; MEVGLKINASE.
CC TRIGRAMS: TIGR00549; mevalon_kin_1.
CC PROSITE: PS00627; GMP_KINASES_ATP_1.
CC TRANSFERASE; Kinase; ATP-binding; Magnesium; Complete proteome.
FT NP_BIND 86 ATP (POTENTIAL).
SQ SEQUENCE 284 AA; 30877 MW; F102C1C71772CA7F CRC64;
Query Match 4.0%; Score 103.5; DB 1; Length 284;

```

```

Best Local Similarity 18.4%; Pred. No. 1.5;
Matches 90; Conservative 61; Mismatches 126; Indels 213; Gaps 20;

QY 3 VVASAPGKVLMTGGYVLEKPNAGVLSTNARFAIYKPIKEVYKPPESMAKMDVLTLS 62
DB 1 MIASAPGKILLEGHAYVGR-----HAAVSAILRCR-----SVKRS 40
QY 63 POLRESMYKLSLNLHLTLOSASASDRNPEVHAIOVAIAAHLATEKDESKLLLOG 122
DB 41 RFLRSSIGESGLDY-----QRHPYVQAVK-----66
QY 123 LDITILGNDFFSYRNOIESAGLPLTPESLGTLPASIFEMAEISGANSKPPVANTGL 182
DB 67 -----RFGELRN-IPGAEIEESB-----IPIGS-----GL 91
QY 183 GSSAAMTAVVAALLHLVGVVDSLPCKEKF--GCSLDYIHMTAQTSHCLAQKVGSGF 241
DB 92 GSSAAVAVATIAL-------NAEFDGDMKEALFQAKOVEIDVGR--ASGI 136
QY 242 DVSCAVVGSQRVAFSEVLSPAQVAVTGLPLNEVIGITLKGW---DNKRTEFSLPLM 298
DB 137 DPFISTFG-----GSMLEPERKVEMPKFEV 163
QY 299 NLFGEPSGGSSTPSVYGAVKWQMSDPEKARENMONSDANLEETKINDLSKLAKDH 358
DB 164 INF-----GSRSTAEMVAKVAELRERHPEVDKIFDAIDALSIE-----202
QY 359 MDVLYRYKSCSVLTSEKWLATEPINEAIKELLEAREAMLRILRONGEASVPI 418
DB 203 -----ASDVSAER-----LEELIALNOSILR-----AIGS- 229
QY 419 EPESOTOLDSTMS--AEGVLLAGVPGAGFDPAIFATILGSGTKLQWASHNVALLV 476
DB 230 NPE-----IDTIALDELMGMNAKITAGGCGCIFGLFKGKPRG-----STIV 273
QY 477 REDPHVCLE 486
DB 274 EPEKEGVIE 283

RESULT 15
LDHA_BRARE
ID LDHA_BRARE STANDARD; PRT; 332 AA.
AC 09FVK5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-lactate dehydrogenase A chain (EC 1.1.1.27) (LDH-A).
GN LDHA.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Osteichthyes; Cypriniformes;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7953;
RN 11
RP SEQUENCE FROM N.A.
RA Tsai S.C.-M., Li J.Y., Mannen H., Li S.S.-L.;
RT "Molecular evolution of vertebrate lactate dehydrogenase isozymes by
RT gene duplication."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: (S)-lactate + NAD(+) -> pyruvate + NADH.
CC -1- PATHWAY: Anaerobic glycolysis; final step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



GenCore version 5.1.4.p5.4578.  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:50:30 ; Search time 39 Seconds  
(without alignments)  
1725.425 Million cell updates/sec

Title: US-09-988-863A-2

Perfect score: 2581

Sequence: 1 MAVVASAPGKMTMGTYLV.....ESGDPRTTCISGVSSIHLE 505

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.GeneSeq.101002.\*

1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2581	100.0	505	23	Arabidopsis thaliana
2	480	18.6	432	22	Protein encoded by
3	475	18.4	432	22	Candida albicans p
4	372.5	14.4	109	21	Pinus radiata phos
5	178.5	6.9	368	22	Enterococcus faeca
6	178.5	6.9	368	22	Mevlonate pathway
7	155	6.0	374	22	Streptomyces sp. C
8	149.5	5.8	359	23	Listeria monocytog
9	147.5	5.7	361	22	Mevlonate pathway
10	142.5	5.5	358	22	S. epidermidis ope

11	142.5	5.5	358	22	AAB60883
12	142.5	5.5	362	23	ABP40510
13	136.5	5.3	358	22	AAB60864
14	135.5	5.2	358	22	AAU37273
15	133.5	5.2	345	22	AAU34358
16	130.5	5.1	358	22	AAB60882
17	123	4.8	378	21	AAAG12901
18	123	4.8	378	21	AAAG50090
19	123	4.8	378	21	AAAG51424
20	123	4.8	380	21	AAAG51423
21	123	4.8	398	21	AAAG50089
22	123	4.8	447	21	AAAG50074
23	123	4.8	451	21	AAAG50073
24	123	4.8	456	21	AAAG50072
25	117	4.5	396	12	AAAR13720
26	116.5	4.5	952	22	AAAT78089
27	115.5	4.5	723	22	AAAB00773
28	115	4.5	432	21	AAAY43633
29	115	4.5	816	23	ABBS3952
30	114	4.4	823	22	ABBT1484
31	112	4.3	287	23	AAO14408
32	112	4.3	338	22	AAAB6178
33	109.5	4.2	369	11	AAAR05738
34	108.5	4.2	1477	14	AAAR41728
35	108.5	4.2	1477	15	AAAR63506
36	108.5	4.2	1477	21	AAAB01848
37	108	4.2	492	22	ABG06982
38	108	4.2	492	22	ABG15375
39	106	4.1	1056	21	AAAG31894
40	106	4.1	1073	21	AAAG31893
41	106	4.1	1087	21	AAAG31892
42	105.5	4.1	308	22	AAAT79592
43	105.5	4.1	931	22	AAAT78649
44	105.5	4.1	949	22	AAAT78649
45	105.5	4.1	949	22	AAAT78649

#### ALIGNMENTS

RESULT 1	
ABBT7505	ABBT7505 standard; Protein; 505 AA.
ID	
XX	ABBT7505;
AC	29-JUL-2002 (first entry)
DT	Arabidopsis thaliana PMVK SEQ ID NO 2.
DE	Thale cress; PMVK, phosphomevalonate kinase; plant; herbicide;
KW	growth regulator; enzyme.
KW	Arabidopsis thaliana.
OS	DE10057755-AL.
PN	23-MAY-2002.
PD	22-NOV-2000; 2000DE-1057755.
XX	22-NOV-2000; 2000DE-1057755.
PF	22-NOV-2000; 2000DE-1057755.
XX	22-NOV-2000; 2000DE-1057755.
PR	(FARB ) BAYER AG.
PA	Weissner R, Lechelt-Kunze C;
XX	WPI; 2002-445360/48.
XX	N-PSDB; ABL60244.
DR	New nucleic acid encoding plant phosphomevalonate kinase, useful for
XX	identifying modulators, potentially useful as herbicides and growth
PT	regulators
PT	

XX Claim 14; Page 13-14; 18pp; German.  
 PS  
 CC The invention relates to a nucleic acid (1, AB160244) that encodes a  
 CC plant phosphomemalonate kinase (PMVK, AB177505), excluding the known  
 CC fully defined partial sequences AB160245-AB160247. Plant PMVK (1),  
 CC constructs and host cells that contain (1) are used to identify agents  
 CC that bind to and/or modulate activity of PMVK, potentially useful as  
 CC herbicides and growth regulators. (1) is also used for recombinant  
 CC production of PMVK.  
 CC  
 XX  
 XX Sequence 505 AA;  
 SO  
 Query Match 100.0%; Score 2581; DB 23; Length 505;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-236;  
 Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAVVASAPGKYLMTGGVLYLEKPNAGLYLSTNARFAYIVKPINEVVPESAMKWTQKL 60  
 DB 1 MAVVASAPGKYLMTGGVLYLEKPNAGLYLSTNARFAYIVKPINEVVPESAMKWTQKL 60  
 QY 61 TSPOLSRSMYKLSLNLTLQSVASDSRNPFEVHAIOYATAAHLATEKDESKLHKL 120  
 DB 61 TSPOLSRSMYKLSLNLTLQSVASDSRNPFEVHAIOYATAAHLATEKDESKLHKL 120  
 QY 121 QGLDITILGSNDYSTRNOIESAGLPLTPESLGTLPAPASITFNAAESNGANSKPEVAKT 180  
 DB 121 QGLDITILGSNDYSTRNOIESAGLPLTPESLGTLPAPASITFNAAESNGANSKPEVAKT 180  
 QY 181 GLSSAAMTAVVAALLHYLGVDLSDPCKEGKFCGSDLDVTHMAQVSHCLAGKVGSG 240  
 DB 181 GLSSAAMTAVVAALLHYLGVDLSDPCKEGKFCGSDLDVTHMAQVSHCLAGKVGSG 240  
 QY 241 FDVSCAVYGSORRYRSEPEVLSPFAQVAVTGLPLNEVIGTILKGMKDNKTEFSLPLMLNL 300  
 DB 241 FDVSCAVYGSORRYRSEPEVLSPFAQVAVTGLPLNEVIGTILKGMKDNKTEFSLPLMLNL 300  
 QY 301 FLGEPGSGSSTPSMGAVKWKQMSDPEKARENMQNSDANLELETKINDLSKLAKDHD 360  
 DB 301 FLGEPGSGSSTPSMGAVKWKQMSDPEKARENMQNSDANLELETKINDLSKLAKDHD 360  
 QY 361 VYLAVIKSCSVLTSEKWLHATEPINEAIIKEILEAREAMLRIRILMKOMGEAASVPIEP 420  
 DB 361 VYLAVIKSCSVLTSEKWLHATEPINEAIIKEILEAREAMLRIRILMKOMGEAASVPIEP 420  
 QY 421 ESQTOLDSTMSAGVLAGVPAGGDFAPATLGDGSGTKLTQAMSSHNVALLVREDP 480  
 DB 421 ESQTOLDSTMSAGVLAGVPAGGDFAPATLGDGSGTKLTQAMSSHNVALLVREDP 480  
 QY 481 HGVCLSGDPRPTTCITTSVSSIHLE 505  
 DB 481 HGVCLSGDPRPTTCITTSVSSIHLE 505  
 RESULT 2  
 AAU15094  
 ID AAU15094 standard; Protein; 432 AA.  
 AC AAU15094;  
 XX  
 XX 04-DEC-2001 (first entry)  
 DE Protein encoded by C. albicans essential gene CayMR220W (ERG8).  
 XX  
 XX Gene identification: essential gene; GRACE; pathogenic fungus;  
 KM gene replacement and conditional expression; fungal infection.  
 XX  
 OS Candida albicans.  
 XX  
 XX WO200160975-A2.  
 XX  
 PD 23-AUG-2001.  
 XX

PF 20-FEB-2001; 2001WO-US05551.  
 XX  
 XX 18-FEB-2000; 2000US-0183534.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Roemer T, Jiang B, Boone C, Bussey H;  
 XX  
 XX WPI; 2001-489080/53.  
 DR N-PSDB; AAS23422.  
 XX  
 PT Identifying genes essential to fungal metabolisms and identifying  
 PT potential therapeutic agents that target these genes -  
 PS  
 PS Claim 43; Page 226-227; 324pp; English.  
 XX  
 CC The present invention relates to novel methods for constructing fungal  
 CC strains useful for identification and validation of gene products as  
 CC targets for therapeutic agents, for creating a collection of identified  
 CC essential genes, and screening assays for the discovery of new drugs.  
 CC The invention provides the GRACE (gene replacement and conditional  
 CC expression) method for the construction of mutant organisms referred to  
 CC as GRACE strains of the organism. The invention can be applied to any  
 CC organism, particularly a pathogenic fungus e.g. Candida albicans,  
 CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are  
 CC useful to identify agents that may be used in the treatment of fungal  
 CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans  
 CC essential genes.  
 CC  
 XX  
 SO Sequence 432 AA;  
 Query Match 18.6%; Score 480; DB 22; Length 432;  
 Best Local Similarity 28.5%; Pred. No. 1.4e-76;  
 Matches 144; Conservative 78; Mismatches 183; Indels 100; Gaps 15;

QY 1 MAVVASAPGKYLMTGGVLYLEKPNAGLYLSTNARFAYIVKPINEVVPESAMKWTQKL 60  
 DB 1 MSAKAPGAKFLAGGVLVLEPYDAYVTALSSMHAVITPKTSLE-----SRKI 53  
 QY 61 TSPOLSR-ESMYKLSLNLTLQSVASDSRNPFEVHAIOYATAAHLATEKDESKLHKL 119  
 DB 54 SSPFANGEMVYHLSN--TEKPEVGSRIKPFLEATIFIVLAVIOPTLEARD----- 103  
 QY 120 QGLDITILGSNDYSTRNOIESAGLPLTPESLGTLPAPASITFNAAESNG-----AN 172  
 DB 104 ---LEIIYSDPGYHSQED-----TETKTSNGEKTEFLYHSR 137  
 QY 173 SKPEVARTGSSAAMTAVVAALL-HYL-GVVDLSDPCKEGKFCGSDLDVTHMAQVSH 230  
 DB 138 ALTEVEKTLGSSAGLVAVATSLSHFIPNVIS-----TKDILHVAQIAH 185  
 QY 231 CLAQKVGSGFDVSCAVYGSORRYRSEPEVLSPFAQVAVTGLPLNEVIGTILK---GKMD 286  
 DB 186 CYACKTKIGSGFDVATATIGSYVRRFOPALINDVGYLESOP--EKPFTELKULIASWE 243  
 QY 287 NKRTFSLPLMLNLFLGEPGSGSSTPSMGAVKWKQMSDPEKARENMQNSDANLELET 346  
 DB 244 FKHERCTLPHGKILKMGDV--KGSETPKLYSRVQMKKEESSVYVDQNSANLPMK 302  
 QY 347 KLNDLSKLAKDHDVYLVRIKSCSVLTSEKWLHATEPINEAIIKEILEAREAMLRIRIL 406  
 DB 303 ELREKREKYSDDPETYIKED-----HSEPLIYAL-----KNIRKG 339  
 QY 407 MRQMGASVPIEPESQTLSDSTMSAGVLAGVPAGGDFAPATLGDGSGTKLTQAM 466  
 DB 340 LQALTQKSEVIEPDVQTLDRQCEIPGCGVVPAGGVDALAVLLEHQVNFQOKT 399  
 QY 467 SS-----HNVALLVREDPHGVCLE 486  
 DB 400 LENDPYFHNVYVVDLEBQTEGVLEE 424

RESULT 3

AA72679  
ID AA72679 standard; Protein: 432 AA.  
XX  
AC AA72679;  
XX  
DT 31-MAY-2001 (first entry)  
XX  
DE Candida albicans phosphomevalonate kinase (PMK; ERG8) protein.  
XX  
KW Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;  
KM Infection.  
XX  
OS Candida albicans.  
XX  
FH Key location/Qualifiers  
FT Misc-difference 244 /note="Encoded by TTC of the sequences shown in  
FT AAD02791 and AAD02792"  
XX  
PN WO200114533-A2.  
XX  
PD 01-MAR-2001.  
XX  
PF 15-AUG-2000; 2000WO-GB03100.  
XX  
PR 21-AUG-1999; 99GB-0019766.  
XX  
PA (ASTR) ASTRAZENECA AB.  
PA (ASTR) ASTRAZENECA UK LTD.  
PI Rosamond JDC, Schnell NF;  
PI  
DR WPI: 2001-218441/22.  
DR N-PSDB: AAD02791, AAD02792.  
XX  
PT New polypeptides and polynucleotides (ERG8) from Candida albicans,  
PT useful in assays for identifying inhibitors of phosphomevalonate kinase  
PT activity and as reagents for diagnosing C. albicans infection -  
XX  
PS Claim 1; Page 27-28; 29pp; English.  
XX  
CC The present sequence is phosphomevalonate kinase (PMK; ERG8) protein  
CC from Candida albicans. The ERG8 protein is useful in an assay for  
CC identifying compounds that inhibit phosphomevalonate kinase (PMK)  
CC activity. These inhibitors are useful as anti-fungal agents. The ERG8  
CC DNA and protein are also useful as reagents for diagnosing C. albicans  
CC infection.  
XX  
SQ Sequence 432 AA;  
Query Match 18.4%; Score 475; DB 22; Length 432;  
Best Local Similarity 28.3%; Pred. No. 4.3e-36;  
Matches 143; Conservative 78; Mismatches 184; Indels 100; Gaps 15;  
QY 1 MAVVASAGKVLMTGTYLLEKPNAGLYLSTNARFYATKPNIEKPKESMAKWTDVLT. 60  
DB 1 MSKAFSAGKAFVLAGYVLEPIYDAYVTALSSRMHAYITPGTSLKE-----SRKI 53  
QY 61 TSPQLSR-ESMYKLSLNHLTQVSASDSRNPFEHAIOYATAAHLATEKRESLKLIL 119  
DB 54 SSPQANKEWEYHISN--TEKPREVQSRINPLEATITFIYAIQPTFAFD----- 103  
QY 120 LQGLDITLIGSNDFYSYNRQIESAGLPLTPESLGTLPASTITFMAAESNG-----AN 172  
DB 104 ---LEITIIYSDPGYHSD-----TEKRTSSNGEKTELYHSR 137  
QY 173 SKPEVAKTGLSSAAMTTAVVAALL-HYL-GVVDLSDPCKEKGFCGSDLVITIMIAGTSH 230  
DB 138 ATTEVEKTGIGSSAGLVSVATSLSHFIPNVIS-----TNKDIIHNAVQIAH 185  
QY 231 CLAQKVGSGPFCVCAVVGSGQRVYFESPEVLSFAQVAVYTGPLNVEYITILK----GKWD 286  
DB 186 CYAQRKISSGFDVAVATYGLVYRRFOPALINDVQVLESDF--EKFPTEKLKLESNWE 243

QY 287 NKRTFESLPPLMNLFLGEPGSGSSTPSVGAVKMKOMSDPEKARENMONLSDANLELET 346  
DB 244 EKHERCTLPYGIKLMGDV-KGSETPKIVSRVYLQMKKEPSSVYDQLSANLQPMK 302  
QY 347 KLNDLSKLAKDMVDYLYRYIKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRIL 406  
DB 303 ELREKREKVDSPETIYKELD-----HSVEPLTVAI-----KNIRKG 339  
QY 407 MRQGEAASVPIEPESQQLDSTMSAGCVLLAGVPGAGGFATPATITLGDSTLQDAM 466  
DB 340 LQALTQKSEVPIEPDVOQLDRCCOIEPCVGVGAGYDAIVLLENQVGNFKQKT 399  
QY 467 SS-----HNVLLALVREDPHCYCLE 486  
DB 400 LENPDYFHNVYWDLEBOTEGVLEE 424  
RESULT 4  
AAB18130  
ID AAB18130 standard; Protein: 109 AA.  
XX  
AC AAB18130;  
XX  
DT 08-NOV-2000 (first entry)  
XX  
DE Pinus radiata phosphomevalonate kinase protein SEQ ID NO:291.  
XX  
KW Eucalyptus grandis; Pinus radiata; modification; isoprenoid; plant;  
KW metabolism; isoprenoid biosynthetic pathway; terpenoid; steroid;  
KW genome mapping; physical mapping; positional cloning; forestry;  
KW agriculture; medicine; fermentation; plant development; pest resistance;  
KW pinene; myrcene; Monterey pine.  
XX  
OS Pinus radiata.  
XX  
PN WO200036081-A2.  
XX  
PD 22-JUN-2000.  
XX  
PF 16-DEC-1999; 99WO-NZ00219.  
XX  
PR 17-DEC-1998; 98US-0215504.  
PR 29-JUL-1999; 99US-0146441.  
XX  
PA (GENE-) GENESTIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
PI Havukkala IJ;  
XX  
DR WPI: 2000-431575/37.  
DR N-PSDB: AAA69566.  
XX  
PT New plant polynucleotides encoding polypeptides involved in the  
PT production and modification of isoprenoids, useful in forestry and  
PT agriculture for manipulation of isoprenoid metabolism -  
XX  
PS Claim 26; Page 159; 164pp; English.  
XX  
CC The present invention describes plant polynucleotides encoding  
CC polypeptides involved in the production and modification of isoprenoids,  
CC such as terpenoid and steroid compounds. The polynucleotides are used  
CC in genome mapping, in physical mapping and in positional cloning of  
CC genes. The polynucleotides and polypeptides are useful in forestry and  
CC agriculture for manipulation of isoprenoid metabolism, in medicine for  
CC therapeutic effects, including direct application in diseased organisms  
CC or indirect application by transgenic organisms and in fermentation and  
CC chemical processing industries involving isoprenoids. In plant  
CC applications, manipulating isoprenoid pathways or isoprenoid composition  
CC may, for example, affect plant development, pest resistance, and the  
CC value of extractives (e.g. pinene and myrcene). The ubiquitous and  
CC varied roles of isoprenoids make the polynucleotides attractive targets  
CC for biotechnical applications in a variety of fields. AAA69527 to

CC AAA69690 and AAB18004 to AAB18143 represent *Escherichia coli* grandis and *Pinus*  
CC radiata polynucleotides and proteins used in the exemplification of the  
CC present invention.

XX Sequence 109 AA;

Query Match 14.4%; Score 372.5; DB 21; Length 109;  
Best Local Similarity 68.6%; Pred. No. 2.5e-27;  
Matches 72; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

QY 1 MAVVASAPGKVTMTGTYLLEKPNAGLVLTSTNAREFAIVKPIINEVPEKSWAKWTQVTL 60  
Db 1 MAVVASAPGKVTMTGTYLLEKPNAGLVLTSTNAREFAIVKPIINEVPEKSWAKWTQVTL 60

QY 61 TSPOLRESMYKLSLNHLTLOSASDSR-NPEVEHAIQYAIYAAA 104  
Db 61 TSPOLAKEAYKLSLKTLSLQNVASSSSNGNPFVEQAVQFVAAA 105

RESULT 5  
AAU35063  
ID AAU35063 standard; protein; 368 AA.

XX AAU35063;

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #350.

KW Antisense; prokaryotic cellular proliferation protein;

KM antibiotic; antibacterial; drug design.

OS Enterococcus faecalis.

PN MO200170955-A2.

PD 27-SEP-2001.

PE 21-MAR-2001; 2001WO-0509180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207272P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;

XX Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX N-PSDB: AAS59922.

XX Example 3; Seq ID No 10656; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 368 AA;

Query Match 6.9%; Score 178.5; DB 22; Length 368;  
Best Local Similarity 22.3%; Pred. No. 5.6e-08;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;

QY 3 VVASAPGKVTMTGTYLLEKPNAGLVLTSTNAREFAIVKPIINEVPEKSWAKWTQVTL 62  
Db 2 IEVTPGKFLFAGYAVEGHPAIIAIVDOFVVEETFEEDSISQASQYSLPIRMT- 60

QY 63 POLRESMYKLSLNHLTLOSASDSR-NPEVEHAIQYAIYAAAHLATKOKESLKLLOG 122  
Db 61 ---RRNGELVLDLR-----ENPF-----HYVLAIIHL-TEKYAQEONKEL--- 96

QY 123 LDIITLGSNDPFYSYRNQIESAGLPLFPEISGLTAPFASITFNMAESNGANSKPEVAKTGL 182  
Db 97 -----SFYHLK-----VTSELDSSNG-----RKYGL 117

QY 183 GSSAAMTTAAVVAAL-LHYLGVDLSDPCKEKGFGCSDLDTVIMIAOTSHCLAQKVGSG 240  
Db 118 GSSGAVTVGTYKALINITYDGL-----ENEIFRLSALHIAVQGN-GSC 161

QY 241 FDSGCAVYGSGRYRSEPEVLSEFAQVAVTGLPLNEVIGTILKGMQNKRT-ESLPLPM 299  
Db 162 GDIAASCYG--WIAFTSFHDWVNOKVT---TETLTDLADWDPELMTEPLKVPQOLR 215

QY 300 LFLGPGSGGSPSPSMGAVKKNQMSPEKAREKNQMLSDANLELETKLNDLSKLARDHW 359  
Db 216 LLIIGTGS-PASTSDLDVDR---HQSKEKQAATEQFLMSRLCVERMINGFN----- 264

QY 360 DYLIRVIRKSCSVLTSEKWLHATEPINEAIIKELLEAREMLRIRIIRMGGEAASVPIE 419  
Db 265 -----TGKIS-VIOKQITKNRQ-----LIAELSSLTGVVIE 294

QY 420 PESOTQLDSTMSAEGVLGAVPGAGGEPDAIFATLGDG--TKLTQWSSHNVLAL 474  
Db 295 TEALKNLCDLAEISTGA--AKSSGAGGDCGIVIFROKSGILPLMTAMERDGIPL 348

RESULT 6

AAAB60859  
ID AAB60859 standard; protein; 368 AA.

XX AAB60859;

XX 30-MAR-2001 (first entry)

XX Mevalonate pathway protein #3.

XX Mevalonate pathway; disease; infection.

XX Enterococcus faecalis.

XX WO200078935-A1.

XX 28-DEC-2000.

XX 22-JUN-2000; 2000WO-0517262.

XX 22-JUN-1999; 99US-0140519.

XX 02-AUG-1999; 99US-0146682.

XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.



PI Brown JR, Gwynn M, Mathie TB, Myers JE, Trainl CM, Van Horn S;  
PI Wilding EI;  
XX  
DR WPI: 2001-071392/08.  
XX  
PT New isolated mevalonate pathway gene polynucleotide derived from  
PT bacterium is useful for treatment of bacterial infection -  
XX  
PS Claim 20; Page 14; 158pp; English.  
XX  
CC The present invention relates to an isolated mevalonate pathway gene  
CC derived from a bacterium from clade of Class II of the  
CC phylogenetic tree referred to in the specification. The invention may be  
CC used for treatment of disease related to bacterial infection, e.g.  
CC conjunctivitis, pneumonia, bacteremia and meningitis.  
XX  
SQ Sequence 368 AA;  
  
Query Match 6.9%; Score 178.5; DB 22; Length 368;  
Best Local Similarity 22.3%; Pred. No. 5.6e-08;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;  
  
QY 3 VVASAPGVMTGCVLVEKPNAGLVTSTNARFAIVPPIEVEKPEPMAMKMDVKLTLS 62  
DB 2 IEVTPGKFLTAGEAAYVPGHPALIVAVDQFVTVEETDEGSIQAOYSLEPIRMT 60  
QY 63 POLRESMYKLSLNHLTLQSVASDSRNPFEHAIOYIAAHLATEKDKSLHKLILQ 122  
DB 61 ---RRNGELVDIR-----ENPF-----HYLAAILH-TEKXAQEQNKEL--- 96  
QY 123 LDITILGSDNFYSYRNQIESAGLPLTPESLGTLPAPASITFNMAESGANSKPEVAKTGL 182  
DB 97 -----SEYHLK-----VTSELDSSNG-----RKYGL 117  
  
QY 183 GSSAAMTAVVAAL--LHYLGVDLSDPCKEGKFGCSLDVYIHMAQSHCLAQKVGSG 240  
DB 118 GSSGAVYVGTAKALNIEFDLGL-----ENEIIFKLSALAHAVQGN-GSC 161  
QY 241 FDVCAYVGSORVYRFSPEVLSEFAOVAVTGLPLNEVIGITILKGKMDNKRTEFSPLPMN 299  
DB 162 GDIAASCGG--WIAFSFHDHWNOKYT-----TETLDLAMDPELMIFPKYPKDLR 215  
QY 300 LFLGPGSGSSTPSMGAVKRWQMSDEKAREMNQNSDANLETKLNDLSKLAKDHW 359  
DB 216 LLIGWTGS-PASTSDLVDRV---HOSKEKQAAVQOFLMKSRCLVETWINGFN----- 264  
QY 360 DYLRLVTKSCSVLSEKVNLAITEPINEAIIKELIAREAMRIRILMRQGEAASVPIE 419  
DB 265 -----TGKIS-VIOKQITRNO-----LLAEILSLTGVVIE 294  
  
QY 420 PESQTOLDSTMSAEGVLLAGVPGAGFDALFAITLGDGSG-TKLTOAMSSHNVAL 474  
DB 295 TEALKNLCDLAEITYTGA--AKSSGAGGDCGIVIFRQKSGILPLMTANMEKQGITPL 348  
  
RESULT 7  
AAB9728  
ID AAB9728 standard; Protein: 374 AA.  
XX  
AC AAB9728;  
XX  
DT 10-SEP-2001 (first entry)  
XX  
DE Streptomycetes sp. CL190 mevalonate pathway orfC protein SEQ ID NO:10.  
XX  
KW Streptomycetes sp. CL190; mevalonate pathway; actinomycete; cardiant;  
KW isoprenoid compound; osteopathic; cytoslatic; ubiquitons; vitamin K2;  
KW carotenoid; heart disease; osteoporosis; cancer; drug; health food.  
XX  
OS Streptomycetes sp. CL190.  
XX  
XX MO200142476-A1.  
XX  
XX

PD 14-JUN-2001.  
XX  
PF 06-DEC-2000; 2000WO-JP08620.  
XX  
XX 08-DEC-1999; 99JP-0348375.  
XX  
PR (SETO/) SETO H.  
XX (KUZU/) KUZUYAMA T.  
XX  
PI Seto H, Kuzuyama T, Takahashi S, Takagi M;  
XX  
DR WPI: 2001-381696/40.  
XX  
PS N-PSDB: AAH44046.  
XX  
PT Actinomycetes-originated genes of enzymes participating in mevalonate  
PT pathway, applicable in producing e.g. ubiquitons, vitamin K2 and  
PT carotenoids for treatment of heart diseases, osteoporosis and cancer in  
PT drug and health food -  
XX  
PS Claim 6; Page 62-64; 75pp; Japanese.  
XX  
CC The sequence given in AAH44043 represents a DNA sequence isolated from  
CC Streptomycetes sp. CL190, containing a 6798 base pairs (S1), which encodes  
CC the whole enzyme necessary for functioning the mevalonate pathway. The  
CC sequence encodes protein sequences, designated orfA to E and hmgR, which  
CC are used in the mevalonate pathway. The proteins and polynucleotide  
CC sequences encoding them have cardiant, osteopathic and cytoslatic  
CC activities. The genes are applicable in producing e.g. ubiquitons,  
CC vitamin K2 and carotenoids which can be used in the treatment of heart  
CC diseases, osteoporosis and cancer in drugs and health foods. The present  
CC sequence represents the orfC protein from the present invention.  
XX  
SQ Sequence 374 AA;  
  
Query Match 6.0%; Score 155; DB 22; Length 374;  
Best Local Similarity 19.7%; Pred. No. 1e-05;  
Matches 97; Conservative 68; Mismatches 172; Indels 156; Gaps 19;  
  
QY 3 VVASAPGVMTGCVLVEKPNAGLVTSTNARFAIVP-----INEEVPESM 51  
DB 8 IYRHAAPGKFLTAGEAAYVDPGHPALIVAVDRHISTVSDADADGAAVYISSDGPVAV 67  
QY 52 AMKWDVRLTSPOLRESMYKLSLNHLTLQSVASDSRNPFEHAIOYIAAHLATEKD 111  
DB 68 GWRWHDGRLV-----VNDP--DDQQAASALAHV----- 95  
QY 112 KESLKLILQGLDITILGSDNFYSYRNQIESAGLPLTPESLGTLPAPASITFNMAESNGA 171  
DB 96 -----SAIEYVGRLLGERGOKVPALTLTSSRLHEDG-- 127  
  
QY 172 NSKPEVAKTGLSSAAMTAVVAALHLYGV-VYLDSDPCKEGKFGCSLDVYIHMAQSH 230  
DB 128 -----RKFGLSSGAVVAVVAALAAAFGLSTLDEFRILAMLATAYELD----- 171  
QY 231 CLAQKVGSGFDVSCAVYGSORVYRFSPEVLSEFAOVAVTGLPLNEVIGITILKGKMDNKR 290  
DB 172 -----PKSGGDLAASWTGG--WIAVQAPDRAF-----YLDLARRGVDRTLKAPPGHSV 220  
QY 291 EFSLEPLNMLFL-----GEPGSGSSTPSMGAVKRWQMSDEKAREMNQNSDANLEFL 345  
DB 221 R-RLPAPKGLTLEVGWTEGP--ASTASLVS-----DLHRRTRWGSASQOREVE 265  
QY 346 TKLNDLSKLAKDHWVYLRVTKSCSVLSEKVNLAITEPINEAIIKELIAREAMLRIRI 405  
DB 266 T-----TDDCVRSAYTALESGB--DTSLHETRRARQELAR--- 299  
  
QY 406 LMRQGEAASVPIEESQTLDDSTMSAEGVLLAGVP-GAGGFDALFAITLGDGSGTKLT- 463  
DB 300 -----LDDEVGIGITFPKLTALCD--AAEAVGAAKAPGAGGDCGIALDAEASRDLTH 352  
QY 464 --QAMSSHNVAL 474  
DB 353 VRORWETAGVLP 365  
DB

RESULT 8  
 ID ABB49817 standard; protein; 359 AA.  
 AC ABB49817;  
 DT 05-FEB-2002 (first entry)  
 DE Listeria monocytogenes protein #2521.  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 OS Listeria monocytogenes.  
 PN W0200177335-A2.  
 PD 18-OCT-2001.  
 PF 11-APR-2001; 2001WO-FR01118.  
 PR 11-APR-2000; 2000FR-0004629.  
 PA (INSP) INST PASTEUR.  
 PI Buchrieser C, Franjeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dusumger O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Meduenio E, De Pablos B, Wehlend J, Kaerst U, Ertian K, Hauf J;  
 PI Rose M, Voss H;  
 DR WPI; 2002-010914/01.  
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides  
 PS Claim 6; SEQ ID NO 2522; 192pp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 XX it are useful for selecting probes and primers for detecting genes in L.  
 XX monocytogenes and related organisms, and for studying genetic  
 XX polymorphisms and other genomes. The present sequence is a protein  
 XX encoded by the genome sequence of the present invention. Proteins  
 XX expressed from the genome sequence are useful for raising specific  
 XX antibodies, identification of L. monocytogenes and related organisms, and  
 XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 XX B12. The genome sequence and proteins encoded by it are also useful for  
 XX selecting compounds that regulate gene expression and cell replication  
 XX and modulate L. monocytogenes-related diseases. In addition, the genome  
 XX sequence and proteins encoded by it are useful in pharmaceutical and  
 XX vaccines compositions for the treatment or prevention of infections by L.  
 XX monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 SQ Sequence 359 AA;  
 Query Match 5.8%; Score 149.5; DB 23; Length 359;  
 Best Local Similarity 20.5%; Pred. No. 3.1e-05;  
 Matches 99; Conservative 60; Mismatches 146; Indels 177; Gaps 20;

QY 8 PGKYLMTGGLVLEKPNAGLVLTNARFYAVK-----PINEEVKP 48  
 DB 13 PGKYLMTGGLVLEKPNAGLVLTNARFYAVK-----PINEEVKP 48  
 DB 13 PGKYLMTGGLVLEKPNAGLVLTNARFYAVK-----PINEEVKP 48

QY 49 ESNAMKWTDVKLPSPOLRESMYKLSLNHLTLOSASDSRNPFEHAIOYAIAMHAT 108  
 DB 72 DGEHWTF-----AEAINIAT 87  
 QY 109 EKDKESEHKLLQGLDITILGSDNFYRNOIESAGLPTPEISGTLAPASTFMAES 168  
 DB 88 -----TFLSSEGIETLPVKM-----ITETLIDQ 111  
 QY 169 NGANSKPEVAKTGLSSAAMTAVAVALL--HYGVVDLSDPCKEKGSGDLDVIMIA 226  
 DB 112 SG-----AKYGLSSAAMTAVAVINMLMKTFEIMLK-----KLA 149  
 QY 227 QTSCHLAQKAVGSGEDVSCAVYSQRYRPSPEVLSAQVAVTGLPNEVIGTILGKMD 286  
 DB 150 ALSHLVYQNGSGCDIASCMGWIAVTFPDQEWVKH-RLAYKSL-----FTMKPEPW 202  
 QY 287 NKRTFSLPPLMNFLEGPSSGSSPTSMGAVYKKWQMSDEKAREMONLSANLEET 346  
 DB 203 MLOETLEDEPVPTFSVMTGT-PVSTGLVSQIHAFKQED-----SKNYQHL-----T 250  
 QY 347 KLNDLSKLARDHMDVLRVIRKSCSVLTSEKWLHATEPINEAIKELLEREAALRIIL 406  
 DB 251 RNNIEMK-----QIIQAFH--TKDELYLSAIKENRRIQEL----- 285  
 QY 407 MRONGEASVPPEPESOTQLIDSTMSAEVLLAG-VGAGGFDAIFAIT-AGDSGTILQ 464  
 DB 286 ---GTRAGVNIETSLKELEAD---SAENMGAGKSSGSGDGLAFSKTKELAERLVN 338  
 QY 465 AW 466  
 DB 339 EW 340  
 RESULT 9  
 ID AAB60881  
 AC AAB60881;  
 DT 30-MAR-2001 (first entry)  
 DE Mevalonate pathway protein #3.  
 KW Mevalonate pathway; disease; infection.  
 OS Enterococcus faecium.  
 PN W0200078935-A1.  
 PD 28-DEC-2000.  
 PF 22-JUN-2000; 2000WO-US17262.  
 PR 22-JUN-1999; 99US-0140519.  
 PR 02-AUG-1999; 99US-0146682.  
 PA (SMIR) SMITHKLINE BEECHAM CORP.  
 PA (SMIR) SMITHKLINE BEECHAM PLC.  
 PI Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;  
 PI Wilding EI;  
 DR WPI; 2001-071392/08.  
 PT New isolated mevalonate pathway gene polynucleotide derived from  
 PT bacterium is useful for treatment of bacterial infection -  
 PS Claim 20; Page 33; 158pp; English.  
 XX The present invention relates to an isolated mevalonate pathway gene  
 XX derived from a bacterium from clade II of the  
 XX phylogenetic tree referred to in the specification. The invention may be  
 XX used for treatment of disease related to bacterial infection, e.g.

CC conjunctivitis, pneumonia, bacteremia and meningitis.  
XX Sequence 361 AA;

Query Match 5.7%; Score 147.5; DB 22; Length 361;  
Best Local Similarity 21.2%; Pred. No. 4.9e-05;  
Matches 102; Conservative 64; Mismatches 171; Indels 145; Gaps 19;

QY 3 VVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIYKPIINEEVPKPMWAKMTDVLTLS 62  
DB 2 IEVSAPKLIYAGEFYAVVEIGHRAVIAVQFVTVYESARKGSIQSAQYSGMPVWMT- 60  
QY 63 POLSRESMYKLSLNHLTQSVASDSRNPEVHAIOYAIAMHLATEKDESLKLLQ 122  
DB 61 ---RNGEVLDIR-----ENPF-----HYLAIRL-TEKVAQEKNIL- 96  
QY 123 LDTITLISNDFYSYRNOIESAGLPLTPESLGTIAPFASITFMAESGANSKPEVATGL 182  
DB 97 -----SFYDLK-----VTSLEDSNG-----RKYGL 117  
QY 183 GSSAAMTTAVVAALHLYGVVLDSDPCKEGKFGCSDLDVIMHTAQTSHCLAQKVGSGFD 242  
DB 118 GSSGAVVATVAKALNVYAL-----NLSOLEIFRIKIALAN--LAVDNGSCGD 163  
QY 243 VSCAVYGSQRYVRESPEVLSFAQVAVTGLPLNEVYIGTLKMDKRTESLPL- 297  
DB 164 IASCVGCG--WIAFSTPDHMLQEOHQHSISELALDWP-----LSIEPLINPED 213  
QY 298 MNFLPEPGSGSSTPMGAVKWKQMSDEPKARENMONSDANLELETKLINDSLAKD 357  
DB 214 LRLIGTGTG--PASTSDLVQV-----HRSRED----- 240  
QY 358 HMDVYLFKIVKSCVLTSEKVNLTATEPINEAI--IKF--LLEAREMLRIRILROMGEA 413  
DB 241 -----KMAVATQLKNTSECVENEMIKGFKENNVTLLIQMTRKNROLHLSAI 288  
QY 414 ASVPIEESOTOLLDSMTSAGEVLLAGVPGAGFDIAITLLGDSG--TKLTOAMSSHNVL 472  
DB 289 TGVVIERPALNKLCLNLAEOYEGA--AKSSGAGGDCGIVIVDOKSGILPLMSAMEKAKEIT 346  
QY 473 AL 474  
DB 347 PL 348

RESULT 10  
AAC81912  
ID AAC81912 standard; Protein: 358 AA.  
XX  
AC AAG81912;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
XX S. epidermidis open reading frame protein sequence SEQ ID NO:918.  
DE  
XX  
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KM vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN W0200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000MO-US30782.  
XX  
PR 09-NOV-1999; 9905-0164258.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
PI Kimerly MJ;  
XX  
DR WPI; 2001-316495/33.

DR N-PSDB; AAH52762.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 273; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAC81454 to AAC83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX  
SQ Sequence 358 AA;  
Query Match 5.5%; Score 142.5; DB 22; Length 358;  
Best Local Similarity 19.8%; Pred. No. 0.00014;  
Matches 94; Conservative 72; Mismatches 167; Indels 141; Gaps 20;

QY 3 VVASAPGKVLMTGGLYLEKPNAGLVLTSTNARFAIYKPIINEEVPKPMWAKMTDVLTLS 62  
DB 2 IQVAPKFLYIAGEFYAVTEPEYKSLIAVNFVATTEASNK-----VEGSIHS 50  
QY 63 POLSRESM-KLSLNHLTQSVASDSRNPEVHAIOYAIAMHLATEKDESLKLLQ 121  
DB 51 KTLHEPVKFDNRNDRISDVQAKO-----LKVVAIAIV----- 87  
QY 122 GLDTITLISNDFYSYRNOIESAGLPLTPESLGTIAPFASITFMAESGANSKPEVAKTG 181  
DB 88 -----FEQVYRSCNNMLKHFL-----TIDSNLADNSG-----QKYG 119  
QY 182 LGSSAAMTTAVVAALHLYGVVLDSDPCKEGKFGCSDLDVIMHTAQTSHCLAQKVGSGF 241  
DB 120 LGSSAAVAVSVKALNFEYVGLTSLNL-----YIYKLVAVIANKLOSLSSCG- 165  
QY 242 DVSCAVY-GSQRYVRESPEVLSFAQVAVTGLPLNEVYIGTLKMDKRTESLPLMN 299  
DB 166 DIAVSVISGWLASTFPHDWK--QOMETS-----VNDVLEKMPGLHLEPLDAPNME 218  
QY 300 LFLGEPGSGSSTPMGAVKWKQMSDEPKARENMONLSANLELETKLINDSLAKDHW 359  
DB 219 VLIGMTGSPASS--PHIVSEVKRLK--SDP-----SFY 247  
QY 360 DVYLRVTKSCSVLTSEKVNLTATEPINEAIKELEREMRLI--RLIMROMGEASVPI 418  
DB 248 GDFLDOSHAC-----VESLQAFRTNKIVOKKIRIRIRIRISMDSEASVEI 295  
QY 419 EPESOTOLLDSMTSAGEVLLAGVPGAGFDIAITLLGDSGTKLTOAMSSHNVL 472  
DB 296 ETDKLRKLCVGEKHGA--SKTSGAGG-----GDCGITIINKVINDKNI 358

RESULT 11  
AAB60883  
ID AAB60883 standard; protein: 358 AA.  
XX  
AC AAB60883;  
XX



Db 170 DIASVSYSGWLAYSTFDHDMWK-QOMETS-----VNDVLEKMPGLHIEPLQAPENME 222  
QY 300 LFLPEPGSGSSTPSMGAVKKQMSDEPKARENMQNSDANLELETNLDSKLAKDH 359  
Db 223 VLIQMTGSPASS-PHLVSEVRRLK-SDP-----SFY 251  
QY 360 DYLVRVYKSCSVLTSEKWLHATEPINEALIKELLEAREAMLRI-RILROMGEAASVPI 418  
Db 252 GDFLDQSHAC-----VESLQAFKTNKIKGQKMRIRKRITIOSMDEASVEI 299  
QY 419 EPESOTQLDSTMSAEGVLLAGVPGAGFDAIFAITLDSCTKLQAMSSHNVL 472  
Db 300 ETDKRLKLDYGEKHGA--SKTSGAGS-----GDCGTTITNKVIKNI 342

## RESULT 13

AAB60864 standard; protein: 358 AA.

AAB60864;

30-MAR-2001 (first entry)

Mevalonate pathway protein #2.

Mevalonate pathway; disease; infection.

Staphylococcus aureus.

W0200078935-A1.

28-DEC-2000.

22-JUN-2000; 2000WO-US17262.

22-JUN-1999; 99US-0140519.

02-AUG-1999; 99US-0146682.

(SMIK ) SMITHKLINE BEECHAM CORP.

(SMIK ) SMITHKLINE BEECHAM PLC.

Brown JR, Gwynn M, Mathie TB, Myers JE, Trainl CM, Van Horn S;  
Wilding EI;

WPI; 2001-071392/08.

New isolated mevalonate pathway gene polynucleotide derived from  
bacterium is useful for treatment of bacterial infection -

Claim 20; Page 17-18; 158pp; English.

The present invention relates to an isolated mevalonate pathway gene  
derived from a bacterium from clade II of the  
phylogenetic tree referred to in the specification. The invention may be  
used for treatment of disease related to bacterial infection, e.g.  
conjunctivitis, pneumonia, bacteremia and meningitis.

SQ Sequence 358 AA;

Query Match 5.3%; Score 136.5; DB 22; Length 358;

Best Local Similarity 21.5%; Pred. No. 0.00054;

Matches 103; Conservative 60; Mismatches 180; Indels 135; Gaps 23;

3 VYAAPGKVLMTGGYLVIEKPNAGLIVSTNAPFAIVPIINEVKKPESMAKMDVKLTS 62

2 IQVAPGKLYIAGEVAIVE-PGYKSVLLADRF-VTATIEAQQ-----YKGT----- 47

63 POLRESNYKTLNHLTIQSVASDS--RNPVEHAIOVAIAAHLATEKDKESLHKIL 119

48 -----IHSKALHNHPYTFSDSDSIYISOPHAKQUNYVTALEIFEQ----- 90

120 LOGIDITILGSDNFYSYRNOIESAGLPLTPESLGTLPASTTFMAESNGANSKPEVAK 179

Db 91 -----YAKSCDIAMKHFHLTIDS-----NLDDSNQ-----HK 117  
QY 180 TGLGSSAMTTAVVAAALHLYGVNLDSPCKEKGCCSDLVYIHIAOTSICLAGKXGS 239  
Db 118 YGLGSSAAVLVSIVKLVNEFYDM-----KLSNL-YIYKLAIVANNKLOSLSGC 164  
QY 240 GFDVCAVY-GSORVVRSPSEVLSPAOYAVTGLPLNEVIGTILKGMKNKRTF-FSLPPL 297  
Db 165 G-DIASVSYSGWLAYSTFDHEWKH-QIEDT-----IYEEVLKMPGLHIEPLQAPEN 216  
QY 298 MNLFLPEPGSGSSTPSMGAVKKQMSDEPKARENMQNSDANLELETNLDSKLAKD 357  
Db 217 MEVLIGMTGSPASS-PHRYSEVRRLK-SDP-----S 245  
QY 358 HMDVLRVYKSCSVLTSEKWLHATEPINEALIKELLEAREAMLRI-RILROMGEAASVPI 417  
Db 246 FYGDFLEDSSHRC---VKK-LIHAFTNNIKGVQKVRQN-----RPIORMDEARVD 294  
QY 418 IEPESOTQLDSTMSAEGVLLAGVPGAGFDAIFAITLGD-SGTRLQAMSSHNVLAL 474  
Db 295 IETEKRLKLDIAEKYHGA--SKTSGAGGDCGTTITNKVDKEXIYDEWTKHGKIKPL 350

## RESULT 14

AAU37273 standard; Protein: 358 AA.

AAU37273;

14-FEB-2002 (first entry)

Staphylococcus aureus cellular proliferation protein #1443.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Staphylococcus aureus.

W0200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI; 2001-611495/70.

N-PSDB; AAS55132.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

Example 3; Seq ID No 12866; 51pp; English.

The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the  
genes, their use in the discovery of novel antibiotics, the essential  
genes themselves and the encoded proteins. The prokaryotes used are  
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
invention is also useful for the identification of potential new targets

for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).

**SQ** Sequence 358 AA;

Query Match	5.28	Score	135.5	DB	22	Length	358
Best Local Similarity	22.68	Pred.	No. 0.00067				
Matches	108	Conservative	58	Mismatches	178	Indels	133
						Gaps	25

```

OY 3 VASAPGVLTMTGGVLIVLEKRNAGVISTNARFAIYVPIINEEVKPESMAMKMDVLTLS 62
Db 2 IOVARPGYLITAGEAVYTE -FGYKSVLIADLR--VVIATIE-----TD----- 42
OY 63 POLSRESWYKLSLNLHLTLOSVASDSRNPVEHAIOVAIAAHLATEKDEKSHKLLLOG 122
Db 43 -----OYKTLITSKALHNPVTFESR--EDSI-VISDPHAKQ----- 77
OY 123 LDITLGSNDYVSVRNOJESA--GLPTLPRESLGLTAPPASTTFMAESNGANSKPEVAKT 180
Db 78 LNYVYTAIEITEQYAKSCDILMKHPHLTIDS-----NLDOSNG-----HKY 118
OY 181 GLGSSAAMTAVVALLHLHYLSDVLDSPCKEGKGFEGCSLDIVIHMAOTSICLAGKVGSG 240
Db 119 GLGSSAALVASYIVKLVNEFYDM-----KLSNL-YTYKLAVIANKKLOSLSSCG.165
OY 241 FDVSCAVY-GSORYVRSPEVLSPQAVAYGTPLNEVYIGTILKKGKMDNKRFE-FSLPPLM 296
Db 166 -DIAVSYSGLMAVSTFDEHWKHK-QIEDT-----IYEEVLINWPLHLEPQAEEM 217
OY 299 NLFLCEPSSGSGSPFMSYGAVKKMQMDSPEKARENMWOLSDANLELETKLNDLSKLAKDH 356
Db 218 EYLIGWIGSPASS-PHFVSEVKRLK-SDP-----SF 244
OY 359 MDVYLRVYKSCSVLTSEKWLHATEPINEALIKELLEAREAMLRIILRMOMGEASVPI 418
Db 247 YGDFLEDSHRC---VEK-LIHAETNNIKGVOKWVRON-----RTIORMDEKATYDI 295
OY 419 EPESOTOLDLSTMASEGVLLGAVPGAGGFDAIPIATLGD-SGTLTGQAMSHNYLAL 474
Db 296 ETEKTYKLCIDIAEKYHGA--SKTSGAGGGDCGITTINKDVDEKVIYDEMTNHGKPL 350

```

RESULT 15  
AAU34358  
ID AAU34358 standard; Protein; 345 AA

DT	14-FEB-2002	(first entry)
XX		
DE	Staphylococcus aureus cellular proliferation protein #634	
XX		
KW	Antisense; prokaryotic cellular proliferation protein;	
KW	antibiotic; antibacterial; drug design.	
XX		
OS	Staphylococcus aureus.	
XX		
PN	WO200170955-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	21-MAR-2001; 2001WO-US09180.	
XX		

PR	21-MAR-2000;	2000US-191078P
PR	23-MAY-2000;	2000US-206648P
PR	26-MAY-2000;	2000US-207612P
PR	23-OCT-2000;	2000US-242578P
PR	27-NOV-2000;	2000US-253625P
PR	22-DEC-2000;	2000US-257931P
PR	16-FEB-2001;	2001US-269308P
XX		
PA	(ELIT-) ELITRA PHARM INC.	

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD,  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI: 2001-611495/70.  
DR N-PSDB: AAS52217.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
PT  
XX Example 3; Seq ID No 5854; 51bp; English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from Wipo at  
CC [ftp.wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

**SQ Sequence 345 AA;**

Query Match	5.2%	Score 133.5	DB 22;	Length 345;
Best Local Similarity	22.8%	Pred. NO. 0.00098;		
Matches 107; Conservative	57;	Mismatches 173;	Indels 133;	Gaps 25

```

QY 7 ARGVLMGGLVLEKPAAGVILSTNMREFVIAIVRPINEVVRPEBMAKMTDYLTPQLS 66
Db 3 APKGTIYAGEVAATE-PEYKSVLLALDRF-V7ATIEE-----TD----- 39
QY 67 RESWKILSNHLTQSVASDSRNPFEVHAIQVIAAAHILATEKDESLHKLQGDIT 126
Db 40 --QYKGTIHSKALHNHPVTERSD--EDSI--V1SDPHAKQ-----LNVY 78
QY 127 ILGSDNDFSYKNQLESA--GLTPIPESLGTIAPPASTIFMAEENSGANSKREVAKTGGS 184
Db 79 VTATIEFQYVAKSCDIAKHFHLTIDS-----NLDDSSNG-----HKYIGGGS 119
QY 185 SAAMT7AAVAALLHLYLGAVDLSDCCKEGKFCGSDLDVYIMHIAIOSHCLAOKVSGGDFVS 244
Db 120 SAALVAVSYIKVLNEFYDM-----KLSNL-YIKYKLAIVANKKLQSLSSCG-DIA 165
QY 245 CAVY-GSORIYRFEPEVLSFNOVAVTGLPIELVEVGTILIKGMWDKKRF-PSLPPLMLML 307
Db 166 VSVSYGWLAVYTFPHEWVKH-QI-IDT-----TVEEYIKMKNPGLHTEPLCAPNNMEFLI 218
QY 303 GEPGSGGSS7PSTMGAVYKWKQMSDPREKARENQMULSDANLELETKLNDLSKLAKDHVDY 362
Db 219 GWTSGRPAS-PHYVSEVVRK-SDP-----SFGGDF 247
QY 363 LRVTKSGVLTSEKVVLTATPEINAIITKELLEAREAMLRITRIMQMGEAASVPIEPES 422

```

Wed May 7 14:13:42 2003

us-09-988-863a-2.rag

Page 11

Db 248 LEDSHRC---VEK-LIHAEKTNNGVOKKVRON-----RILIORMDKEATVDIETEK 296  
QY 423 QTOlldSTMSAEVLLAGVPGAGGEDAIFAITLGD-SGKTLTOAMSSHN 471  
Db 297 LKYLCDIAEKYHGA--SKTSGAGGGDCGTLINKVDKDKIYDEWTKHGI 344

Search completed: April 26, 2003, 12:55:05  
Job time : 42 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:54:20 ; Search time 17 Seconds  
(without alignments)  
874.034 Million cell updates/sec

Title: US-09-988-863A-2

Perfect score: 2581  
Sequence: 1 MAVVASAPGKVLMTGTYLV.....ESGDPPTTCITSGVSIHLE 505

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA.\*  
2: /cgn2\_6/ptodata/1/laa/5A.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/laa/5B.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/laa/6A.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/laa/6B.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/laa/6C.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142.5	5.5	362	4	US-09-134-001C-5355
2	117	4.5	396	6	US-09-134-001C-5355
3	115	4.5	432	4	US-09-306-595C-8
4	108.5	4.2	1477	1	US-08-038-682-4
5	108.5	4.2	1477	1	US-08-302-832-4
6	108.5	4.2	1477	2	US-08-530-198-4
7	108.5	4.2	1477	2	US-08-469-880-4
8	108.5	4.2	1477	2	US-08-728-470-4
9	108.5	4.2	1477	2	US-08-617-697-4
10	108.5	4.2	1477	2	US-08-719-641-4
11	108.5	4.2	1477	4	US-09-206-942-71
12	104.5	4.0	441	4	US-09-651-200-4
13	104.5	4.0	770	1	US-08-445-135-2
14	101.5	3.9	534	4	US-09-651-200-6
15	101.5	3.9	534	4	US-09-651-200-6
16	101.5	3.9	1024	4	US-09-562-737-4
17	101	3.9	1012	2	US-08-475-891A-4
18	101	3.9	1025	2	US-08-567-375-4
19	101	3.9	1025	2	US-08-587-680A-4
20	97	3.8	863	4	US-08-238-303-11
21	95.5	3.7	1294	4	US-08-819-288-3
22	95.5	3.7	1294	4	US-09-400-348-3
23	95.5	3.7	1321	1	US-08-261-822A-3
24	95.5	3.7	1321	5	PCT-US95-07744A-3
25	95.5	3.7	1599	2	US-08-617-697-9
26	95	3.7	1093	4	US-09-315-793-52
27	93.5	3.6	1536	4	US-09-206-942-67

28	93	3.6	608	4	US-08-843-572E-2	Sequence 2, Appli
29	92.5	3.6	640	4	US-09-177-349-4	Sequence 4, Appli
30	91.5	3.5	1013	4	US-09-415-522-8	Sequence 8, Appli
31	91	3.5	500	6	5486473-2	APPLICANT: FUK
32	91	3.5	730	4	US-09-398-865A-2	Sequence 2, Appli
33	91	3.5	730	4	US-09-710-714-2	Sequence 2, Appli
34	90.5	3.5	587	4	US-09-020-465-2	Sequence 2, Appli
35	90.5	3.5	634	2	US-09-020-466-2	Sequence 2, Appli
36	90.5	3.5	634	3	US-09-197-659-2	Sequence 2, Appli
37	90	3.5	1285	1	US-07-582-945-2	Sequence 2, Appli
38	90	3.5	1285	2	US-08-453-141-2	Sequence 2, Appli
39	90	3.5	1285	3	US-08-293-314-2	Sequence 2, Appli
40	90	3.5	2958	4	US-08-894-344C-2	Sequence 2, Appli
41	89.5	3.5	969	1	US-08-365-689-3	Sequence 3, Appli
42	89.5	3.5	969	1	US-07-747-781-3	Sequence 3, Appli
43	89.5	3.5	969	1	US-08-145-138A-3	Sequence 3, Appli
44	89.5	3.5	969	5	PCT-US92-06888-3	Sequence 3, Appli
45	89.5	3.5	969	5	PCT-US93-03027-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-09-134-001C-5355  
Sequence 5355, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
FILE REFERENCE: GTC-007  
CURRENT FILING DATE: 1998-08-13  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5355  
LENGTH: 362  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5355

Query Match 5.5%; Score 142.5; DB 4; Length 362;  
Best Local Similarity 19.8%; Pred. No. 3e-06;  
Matches 94; Conservative 72; Mismatches 167; Indels 141; Gaps 20;

3 VVASAPGKVLMTGTYLVLEKPNAGLVLTSTNAPFAIVKPIDEEVKPE\$M\$WKMTQVLTLS 62  
6 IOVKAPGKLYINGEVAIVPEGKSLIIANVRVTITIE\$SNK-----VGSIIHS 54  
63 POLSRESM-YKLSLHNLTLQSVASDSRNPFEVHAIOVAIAAHLATERKRESLKLILQ 121  
55 KTLHEPVKFDNENRIEISDVQAAKQ-----LKYVVAIEV----- 91  
122 GIDITILGNDFTSTRNQTIESAGLPLTPESLCTLA\$P\$ITFNAESNCANKPEVATG 181  
92 -----PEQVYRSCNMMLKHFHL-----TIDSNLADNSG-----QRYG 123  
132 LGSSAAMTAVVAALHLHYGVVDLSDPCKEKGFGSGDLDVIMIMQTSHCLAQGVGSGF 241  
134 LGSSAALVSVKALNEFYG-LELSNL-----YIKLAVINMKLQSLSSCG- 169  
242 DVSCAVY-GSORRYAFSPSEVL\$FAQVAATGLPLNEVIGTILKGKWNKRT\$-FSLPLMN 299  
170 DI\$V\$V\$G\$M\$AT\$F\$D\$D\$W\$K-QQMEETS-----VNDVLEKNNPGLHIEQLQ\$PENNE 222  
300 LFLG\$V\$G\$G\$S\$T\$P\$M\$G\$V\$K\$W\$M\$S\$D\$P\$K\$A\$R\$E\$N\$M\$N\$D\$A\$N\$E\$L\$E\$T\$Y\$K\$N\$D\$S\$K\$A\$K\$D\$H\$W 359  
223 VLI\$G\$T\$G\$P\$A\$S\$-P\$H\$V\$E\$V\$K\$R\$K-SDP-----SFY 251

QY 360 DYLRVIRKSCSVLTSEKWLHATEPIINEAIKELLEAREMLRI-RILMOMGEAAVPT 418  
 Db 252 GDLIDOSHAC-----VESLIQAFKTNKICGVQKIRINRIRRIISDMONEASVEI 299  
 QY 419 EPESQOQLDSTMSAEGVLLAGVPAGGFAIFAITIGDSGTIKLTQAMSSHNV 472  
 Db 300 ETDKLKMLCQVGEKHGA--SKTSGAG-----GDCGITTINKVIDKNII 342

## RESULT 2

5349058-2  
 ; Patent No. 5349058  
 ; APPLICANT: TANAKA, RICHARD D.; RICCI, BEVERLY S.  
 ; MOSLEY, STEPHEN T.  
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN  
 ; MEVALONATE KINASE  
 ; NUMBER OF SEQUENCES: 2  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/475,577  
 ; FILING DATE: 06-FEB-1990  
 ; SEQ ID NO: 2  
 ; LENGTH: 396  
 5349058-2

Query Match 4.5%; Score 117; DB 6; Length 396;  
 Best Local Similarity 20.2%; Pred. No. 0.0017;  
 Matches 104; Conservative 74; Mismatches 183; Indels 154; Gaps 22;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEEYKPPSMAMKMTDVLT 62  
 Db 6 LTVSAPGKVLHGEHAAVH-GKVALAVSLNRTFLRIQPHSN-----GKVDLSL 53  
 QY 63 POLSRESMYKLSLNHLTLOSVSASDSRNPFVENAIOYAIAMHLATEKDESLKLLQ 122  
 Db 54 PNIGIKRAMOVA-----RLQSLDTS-----FLEOGD-----VTPTSEQVEKLR--- 92  
 QY 123 LDITILGSDNPFYRNQIESAGLP---LTPESIGTLAPPASITITNAESGANSKPEVA- 178  
 Db 93 -----EVAGLPDDCAVTERLAVALA-FYLYLISICKORALPSLDIV 133  
 QY 179 -----KTGLSSNAAMTAVAAALHLYGVNLSDPCKEG---KFGCSLDIVIHMTAQT 228  
 Db 134 WSELPPAGIGSSAAAYVCCLAAALITV--CEETPNPLKDDGCVMKWKEDLELINKAFQ 191  
 QY 229 SHCIAQKVGSGPDVSCAVYGSQRYVRESPEVLSEAOVAVTGPLNEVIGITILGKWDNK 288  
 Db 192 GERIHGN-PSGVNNAVSTWGC-----ALRYHOGKISSLSKSPALQILT----- 234  
 QY 289 RTEISLPLMLNPLFGEKSGSSSTPSMVGAVKKWMSDPEKARENMONLSDANLEETKL 348  
 Db 235 -TNTKVP-----RMTRALVAGVRNRLKFPPEIVAPDLNLSIDALSIECERVL 279  
 QY 349 NDLSKLAKDHWYLRVIRKSCSVLTSEKWLHATEPIINEAIKELLEAREMLRIRILMR 408  
 Db 280 GEMGE-----APAEQYLVLEELIMNQHHLN----- 306  
 QY 409 OMGEAASVPIEPESQOQLDSTMSAEGVLLAGVPAGGFAIFAITIGDSGTIKLTQAMSS 468  
 Db 307 -----ALGVGHASLDQLCQVT--RARG-LHSKILGAGGGCC--GITLLKGLQEPVEAT 356  
 QY 469 HNVIALLVREDPHGVCELSGDPRTTCTTSCVSSIH 503  
 Db 357 KQALTSCGPD-----CLE-----TSIGAPGVSIH 380

## RESULT 3

US-09-306-595C-8  
 ; Sequence 8, Application US/09306595C  
 ; Patent No. 6284506  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HOSHINO, Tatsuo  
 ; APPLICANT: OJIMA, Kazuyuki  
 ; APPLICANT: SETOGUCHI, Yutaka

; TITLE OF INVENTION: ISOPRENOLID PRODUCTION  
 ; FILE REFERENCE: ISOPRENOLID PRODUCTION  
 ; CURRENT APPLICATION NUMBER: US/09/306,595C  
 ; CURRENT FILING DATE: 1999-05-06  
 ; PRIOR APPLICATION NUMBER: 98108210  
 ; PRIOR FILING DATE: 1998-05-06  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 432  
 ; TYPE: PRT  
 ; ORGANISM: Phaffia rhodozyma  
 US-09-306-595C-8

Query Match 4.5%; Score 115; DB 4; Length 432;  
 Best Local Similarity 21.6%; Pred. No. 0.0032;  
 Matches 103; Conservative 64; Mismatches 190; Indels 120; Gaps 19;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEEYKPPSMAMKMTDVLT 62  
 Db 4 LTVSAPGKVLHGEHAAVH-GKVALAVSLNRTFLRIQPHSN-----GKVDLSL 53  
 QY 63 POLSRESMYKLSLNHLTLOSVSASDSRNPFVENAIOYAIAMHLATEKDESLKLLQ 122  
 Db 49 SSLSSTNI-TTSLDNLNFTQSWPVDS-----LPWSLAPDW--TEASTIPESICPTLLA 97  
 QY 122 GLD-TTILGSDN-----FYSRNQIESAGLPITPESIGTLAPPASITITNAESNGA 171  
 Db 98 EIERIAGGGNGGREGKAVATMAFLYLLVLSKGPSEPFEL-----TANSLPMA 148  
 QY 172 NSKREPAKTLGSSAAMTAVAAALHLYGVNLSDPCKCKGCGCSLDIVIHMTAQTSHC 231  
 Db 149 -----GLGSSAALSTIALAVFLHFSHLSPTTTGRESTIPTADTEVIDKWAFLAEK 199  
 QY 232 LAQKVGSGPDVSCAVYGSQRYVRESPEVLSEAOVAVTGPLNEVIGITILGKWDNKRT 291  
 Db 200 VIGHN-PSGIDNAVSTWGC-----AVAKRKRIEKGQ-E 230  
 QY 292 FSLPPLMLN-----FLGEPGSGSSSTPSMVGAVKKWMSDPEK-----ARENMON 336  
 Db 231 GGMKAISFTSIRPLTDSRIGRDTSLVAGVNARLQIEBEVIVPLLEAIOQLADEAIRC 280  
 QY 337 LSDANLEETKINDLSLAKDHWYLRVIRKSCSVLTSEKWLHATEPIINEAIKELLEA 396  
 Db 291 LKQSEMERAVMIRLOMLVSEN-----HAHLAAGVSHSPSIEETIR-IGAD 335  
 QY 397 REAMLRIIRILMOMGEAASVPIEPESQOQLDSTMSAEGVLLAGVPAGG-GF 447  
 Db 336 KPPELRKILGAGGGCAVTLVPDPEFTETLQALMETLVGSSSFAPYIARVGSQVGF 392

## RESULT 4

US-08-038-682-4  
 ; Sequence 4, Application US/08038682  
 ; Patent No. 5549897  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BARENKAMP, STEPHEN J  
 ; APPLICANT: ST. GEME III, JOSEPH W  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
 ; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Matlare, Ltd  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,682  
FILING DATE: 16-MAR-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-293  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-038-682-4

Query Match 4.2%; Score 108.5; DB 1; Length 1477;  
Best Local Similarity 20.3%; Pred. No. 0.13; Mismatches 162; Indels 119; Gaps 21;  
Matches 89; Conservative 68;

QY 11 VLMTGYLVLEKPNV-----GLVLTNARFYAIKPIKEEYKPSMAWKMT-----56  
DB 139 ILDSNGVFLNPGITIGKDAIINTNGFTASTLDSINENIKARNFTEQTKDALEIV 198  
QY 57 -----DVKLTSQLSRESMYKL--SLNHLTQVSASDSRNPVEHAIOVAIA 102  
DB 199 NHGLITVGKDSVNLIGKVKNEGVISVNGSISILAGOKITISDIINP-----TITYSIA 254  
QY 103 AAHLATEKDKESLKLLOGDITILGSNDFYSYRNQIESAGLPLTPESLGLAPFASIT 162  
DB 255 -----APEEAVNLGDIPRAKGINVRAA---TIRNOKLSADSVSKDKSGNIV--LSAK 304  
QY 163 FNAESNGA-NSKPEVAKTG---LGSSAMTAVVAALL-----HYLGVYDLSDPCKE 211  
DB 305 EGAEIGGVISAQNOQAKGKLMITGKVLTKGAVIDLSGEGGETYLG---GDERGE 360  
QY 212 GKFGCSDLDVTHMTAQOT-----SHCLAQKVGSGFDVSCAVYGSORRYRSEPEVLSPAOV 266  
DB 361 GKNG-----IQLAKKTSLEKSGTINVSCKEKGFAI---VWGDIALIDGINAOGSGDI 411  
QY 267 AVTG-----LPLNEVYIGTILKGMW---DNKRTESLPLIMLFLGEPGSGSS 311  
DB 412 AKTGFEVETSGHDLFIKDAIVDA---KEWLDPDNVSINADPLFNNTGTINDEPPTGTG 468  
QY 312 TPSNVAVKKQMSDPEKARENMONLSDANILEFTKLNLD--LSKILAKDHWV-----YL 363  
DB 469 -----EASDPKK-----NSELKTTLTNTTISVLYKNAKMTMINITASRKL 506  
QY 364 RVIKSCSVLTSEKWLHA 381  
DB 507 TVNSSINIGNSHLIHS 524

RESULT 5  
US-08-302-832-4  
Sequence 4, Application US/08302832  
Patent No. 5603938  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Matzare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia

COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/302,832  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US pct/us93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-404  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-302-832-4

Query Match 4.2%; Score 108.5; DB 1; Length 1477;  
Best Local Similarity 20.3%; Pred. No. 0.13; Mismatches 162; Indels 119; Gaps 21;  
Matches 89; Conservative 68;

QY 11 VLMTGYLVLEKPNV-----GLVLTNARFYAIKPIKEEYKPSMAWKMT-----56  
DB 139 ILDSNGVFLNPGITIGKDAIINTNGFTASTLDSINENIKARNFTEQTKDALEIV 198  
QY 57 -----DVKLTSQLSRESMYKL--SLNHLTQVSASDSRNPVEHAIOVAIA 102  
DB 199 NHGLITVGKDSVNLIGKVKNEGVISVNGSISILAGOKITISDIINP-----TITYSIA 254  
QY 103 AAHLATEKDKESLKLLOGDITILGSNDFYSYRNQIESAGLPLTPESLGLAPFASIT 162  
DB 255 -----APEEAVNLGDIPRAKGINVRAA---TIRNOKLSADSVSKDKSGNIV--LSAK 304  
QY 163 FNAESNGA-NSKPEVAKTG---LGSSAMTAVVAALL-----HYLGVYDLSDPCKE 211  
DB 305 EGAEIGGVISAQNOQAKGKLMITGKVLTKGAVIDLSGEGGETYLG---GDERGE 360  
QY 212 GKFGCSDLDVTHMTAQOT-----SHCLAQKVGSGFDVSCAVYGSORRYRSEPEVLSPAOV 266  
DB 361 GKNG-----IQLAKKTSLEKSGTINVSCKEKGFAI---VWGDIALIDGINAOGSGDI 411  
QY 267 AVTG-----LPLNEVYIGTILKGMW---DNKRTESLPLIMLFLGEPGSGSS 311  
DB 412 AKTGFEVETSGHDLFIKDAIVDA---KEWLDPDNVSINADPLFNNTGTINDEPPTGTG 468  
QY 312 TPSNVAVKKQMSDPEKARENMONLSDANILEFTKLNLD--LSKILAKDHWV-----YL 363  
DB 469 -----EASDPKK-----NSELKTTLTNTTISVLYKNAKMTMINITASRKL 506  
QY 364 RVIKSCSVLTSEKWLHA 381  
DB 507 TVNSSINIGNSHLIHS 524

RESULT 6  
US-08-530-198-4  
Sequence 4, Application US/08530198  
Patent No. 5869065





TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-617-697-4

Query Match 4.28; Score 108.5; DB 2; Length 1477;  
Best Local Similarity 20.38; Pred. No. 0.13;  
Matches 89; Conservative 68; Mismatches 162; Indels 119; Gaps 21;

QY 11 VLMTGVLVEKPNV-----GLVSTNARFYAIKPIINEEVKPSMAKMT-----56  
DB 139 ILDSNGVFLINPGITIGKDAIINTNGFTASTDISINENKARNFTFEQTKKALAEIV 198  
QY 57 -----DKLTPSPQRESMYKL---SLNHLTLOSASDSRNPVEHAIOYAIA 102  
DB 199 NHGLITVGKDSVNLIGKYNKEGVISVNGSISLLAGOKTITSDIINP-----TITYSIA 254  
QY 103 AAHLATEKDKESLHKLLLOGDITILGSNDFYSRNOIESAGLPLTPESIGTAPFASIT 162  
DB 255 -----APENEAVNLGDIFRAKGNINVRRA---TIRNOGKLSDASVSKDSGNIV--LSAK 304  
QY 163 FNAESNGA-NSKPEVAKTG---LGSSAAMTAVVAALL-----HYLGVDLSDPCKE 211  
DB 305 EGAEIGVISAQNOQAKGKMTTGDKVTLTGAVIDLSEKEGEYTLG---GDERGE 360  
QY 212 GKFGCSDLDVIHMTAQT---SHCLAQKVGSGFDVSCAVYSGQRYVRFSEVLSFAOV 266  
DB 361 GKNG-----IQAKKTSLEKSTINVSKEKGGFAI---VMGDIALIDGINNOGSGDI 411  
QY 267 AVTG-----LPLNEVIGTILGKW---DNKRTESLPLNMLFLGEPGSGSS 311  
DB 412 AKTGFEVETSGHDLFIKDNALVDA---KEWLDPDNVSINAEPLFNNTGINDPPTGTG 468  
QY 312 TPSWAGVAKKQMSDPEKARENMONLSDANIELETKLND--LSKLADHMDV-----YL 363  
DB 469 -----EASDPK-----NSELTTLTNTTISYLNKAMTMINITASRKL 506  
QY 364 RVIKSCSVLTSEKWLHA 381  
DB 507 TVNSSINIGNSHLIHS 524

RESULT 10  
US-08-719-641-4  
Sequence 4, Application US/08719641  
Patent No. 6218141  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Matzare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE: 530  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstreser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-719-641-4

Query Match 4.28; Score 108.5; DB 4; Length 1477;  
Best Local Similarity 20.38; Pred. No. 0.13;  
Matches 89; Conservative 68; Mismatches 162; Indels 119; Gaps 21;

QY 11 VLMTGVLVEKPNV-----GLVSTNARFYAIKPIINEEVKPSMAKMT-----56  
DB 139 ILDSNGVFLINPGITIGKDAIINTNGFTASTDISINENKARNFTFEQTKKALAEIV 198  
QY 57 -----DKLTPSPQRESMYKL---SLNHLTLOSASDSRNPVEHAIOYAIA 102  
DB 199 NHGLITVGKDSVNLIGKYNKEGVISVNGSISLLAGOKTITSDIINP-----TITYSIA 254  
QY 103 AAHLATEKDKESLHKLLLOGDITILGSNDFYSRNOIESAGLPLTPESIGTAPFASIT 162  
DB 255 -----APENEAVNLGDIFRAKGNINVRRA---TIRNOGKLSDASVSKDSGNIV--LSAK 304  
QY 163 FNAESNGA-NSKPEVAKTG---LGSSAAMTAVVAALL-----HYLGVDLSDPCKE 211  
DB 305 EGAEIGVISAQNOQAKGKMTTGDKVTLTGAVIDLSEKEGEYTLG---GDERGE 360  
QY 212 GKFGCSDLDVIHMTAQT---SHCLAQKVGSGFDVSCAVYSGQRYVRFSEVLSFAOV 266  
DB 361 GKNG-----IQAKKTSLEKSTINVSKEKGGFAI---VMGDIALIDGINNOGSGDI 411  
QY 267 AVTG-----LPLNEVIGTILGKW---DNKRTESLPLNMLFLGEPGSGSS 311  
DB 412 AKTGFEVETSGHDLFIKDNALVDA---KEWLDPDNVSINAEPLFNNTGINDPPTGTG 468  
QY 312 TPSWAGVAKKQMSDPEKARENMONLSDANIELETKLND--LSKLADHMDV-----YL 363  
DB 469 -----EASDPK-----NSELTTLTNTTISYLNKAMTMINITASRKL 506  
QY 364 RVIKSCSVLTSEKWLHA 381  
DB 507 TVNSSINIGNSHLIHS 524

RESULT 11  
US-09-206-942-71  
Sequence 71, Application US/09206942  
Patent No. 6432669  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
FILE REFERENCE: 1038-861 MIS-Jb  
CURRENT APPLICATION NUMBER: US/09/206,942  
CURRENT FILING DATE: 1998-12-08  
EARLIER APPLICATION NUMBER: 09/167,568  
EARLIER FILING DATE: 1998-10-07  
NUMBER OF SEQ. ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 71  
LENGTH: 1477  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-206-942-71

Query Match 4.2%; Score 108.5; DB 4; Length 1477;  
Best Local Similarity 20.5%; Pred. No. 0.13;  
Matches 90; Conservative 70; Mismatches 157; Indels 123; Gaps 21;

QY 11 VLMGTVLEKRNA-----GLVSTNARFYAIKPIINEEYKPPSMAMKMT-----56  
DB 139 ILDSNGCVFLNPNGITIGKDAIINTNGFTASTLDISENIKARNFTFEQKDALAEIV 198  
QY 57 -----DVKLTSPQLSRESMYKL---SLNHLTLQSVASDSRRPFEHAIQVIA 102  
DB 199 NHGLITWGKDSVNLIGKAKNEGVISVNGSISLQAGKTTISDIINP---TITYSIA 254  
QY 103 AAHLATEKDESLHKLLOGIDITILGSDNFYSYRNOIESAGLPETESLGLTAPFASIT 162  
DB 255 -----APEHEAVNLGDIFAKGGINVRAA---TIRNOGKLSDSVSKXSGNIV--LSAK 304  
QY 163 FNAESGKA-NSKPEVAKT---LGSSAANTVAVALL-----HYGVYDLSDPCKE 211  
DB 305 EGEAEIGGVISAQNOAKGKLMITGDKVTLTGAVIDLSEKGEGETYLG---GDERGE 360  
QY 212 GKFCGSDLDVTHMIAOTSHCLAQGVSGFDVS-----CAYYGSQRYVRFSPVLSFA 264  
DB 361 GKNG-----IQLAKTSL-----EKGSTINVSKEKGRVIMGDIALDGNINAAQSG 409  
QY 265 QVAVTG-----LPLNEVIGTILGKW---DNKRTFSLPPLMLFLGEPGSGG 309  
DB 410 DIARTGFEVETSGHYLSIDSNAIYKT---KEMLDPPDVTLEADPLKNNTGIDEPFTG 466  
QY 310 SSTSMGVAKKKQMSPEKAREMWNISDANLETKLND--LSKLAKMDV-----361  
DB 467 TG-----BASDPK-----NSELKTTLTNTTISNLYKNAMTNITASR 504  
QY 362 YLRVYKCSVLTSEKWLHA 381  
DB 505 KLTIVNSSINIGSNHLIHS 524

RESULT 12  
US-09-651-200-4  
Sequence 4, Application US/09651200  
Patent No. 6429303  
GENERAL INFORMATION:  
APPLICANT: Green et al  
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
FILE REFERENCE: 15966-562 (CURA-62)  
CURRENT APPLICATION NUMBER: US/09/651.200  
CURRENT FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152283  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/172909  
PRIOR FILING DATE: 1999-12-21  
PRIOR APPLICATION NUMBER: 60/183578  
PRIOR FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 441  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-651-200-4

Query Match 4.0%; Score 104.5; DB 4; Length 441;  
Best Local Similarity 22.3%; Pred. No. 0.042;  
Matches 87; Conservative 45; Mismatches 112; Indels 147; Gaps 22;

QY 209 CKEGFGCSDLDVTHMIAOTSHCLAQKV---GSGFDVSCAVYGSQRYVRFSPVLSFAV 266  
DB 39 CRQPAQSGSLLE-----AQHPGAQOQGAARGHGVITICSSY--QGY-----PEAEVQMD 87  
QY 267 AVTGLPLNEVIGTILGKWKDKRTFESLPPLMLFLGEPG-----SGGS 310  
DB 88 G-QGVPLT---GNVTTSSMANEGCLFDVHSILRVYLGANGYSCLVNRPNVLODPAHSSVT 143  
QY 311 STP---SMGVAKKKQMSPEKAREMWNISDANLE-----TKLNDLSKIAKQHW 359  
DB 144 ITPQSRPTGAAYEVOVPEDPVALVG---TDATLHCSFSPPEGFSLTQNLIL-----W 192  
QY 360 DVYLRVYKCSVLTSEKWLHA-TEPINE-----AIKELLEAREAMLRIR-----404  
DB 193 Q-----LDTQOLVHSTFEGRQGSAYANRRTLLFPDILAQNASLRQVRVAD 241  
QY 405 -----ILMRQGBAA-----SVPIEPESOTQLDST-----430  
DB 242 EGSFTCEFSIRDFGSAVSLQVAAPYSKPSMTLEPNKDLRPGDVTITCSSYRGYPEAEV 301  
QY 431 --MAEGLVLLG-----VYAGGFD--AIFAITLGDSSGKTLQAMSHVALLVR-- 477  
DB 302 FWDQGGVPLTGNVTTSSMANEGCLFDVHSYLRVYLGANGT-----YSCLVANP 350  
QY 478 --EDPHGVCLSESDPRT---TCITSGVS 500  
DB 351 VLQDPAHGSVITIGQPMFPEALMVTGLS 381

RESULT 13  
US-08-445-135-2  
Sequence 2, Application US/08445135  
Patent No. 5658789  
GENERAL INFORMATION:  
APPLICANT: Quaranta, Vito  
APPLICANT: Hormla, Markella  
TITLE OF INVENTION: Promotion of Epithelial Cell Adhesion  
TITLE OF INVENTION: and Hemidesmosome Assembly by a Laminin-Like Molecule  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445.135  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/151,134  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: DESMOS.002CPE2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
IMMEDIATE SOURCE:  
CLONE: 150 KD  
US-08-445-135-2

Query Match 4.08; Score 104.5; DB 1; Length 770;  
Best Local Similarity 19.18; Pred. No. 0.11;  
Matches 98; Conservative 65; Mismatches 196; Indels 153; Gaps 17;

QY 67 RESM--YKSLNHL--TIGSVASDSNRPVFAIQAIAAHATKDESLKLLLG 122  
DB 125 RDSLNDTEALQDRLSVYQDPAAGKATGLNHNESVLAIQ-RQKEMDSLKLYTEH 183  
QY 123 L--DITILGNDFYSTRNOIESAGLPLTPESLGTLLAPFASITFNAESGANSKEPVAK 179  
DB 184 LAAVADASLQTNLSLQMDTSQK-----EYESL-----AALANGRELNDQY 226  
QY 180 TGLGSSAAMTAVVALLHLGLVVDLSDPCKEKGCGSDLDVTHMTQTSCHL----- 232  
DB 227 RELSRSGKAPLYAEAKHAQSLOELAKOLEIKRNTSGESVRCVDAATAYESIIMAI 286  
QY 233 -----AGKYGSGFDVSCAVYSGQRYRFSPEVLSFAQVAVTGLP-----LN 274  
DB 287 RAEDDAKGRADSASE-----SAFOTVAKEDLPRAKTLSSDSEELN 328  
QY 275 EVIGTITLKGMDNKRREFSLPPLMN-----LFLGEP 305  
DB 329 E-----AMTRKRRLQOEINPALNSLQTLKTVSYQKDLLANVAVRNDLRIGQSDI 381  
QY 306 GSGGSSTPSVV---GAVKKQMSDPEKAREWQNL-----SDANLETKLNDLSK 353  
DB 382 DSVVSGAKSVRRKANGTSEVLDSAPSRIWGLRTATCTRHEDPKALIIANNVYKK 441  
QY 354 LAKDMDVLYRVIKSCSVLSEKVVHATEPINEAI-----IKEEL-EAREAM 400  
DB 442 LTKKLPLPVKI-----ESINQQLPLIGNISNVRIEHLITQADAA 484  
QY 401 LRIRILRQMGEAASVPIEPESQTLLDSTMSAGVLLAGVPGAGFDPAITLGDST 460  
DB 485 NKVAIPRFKNGKSGVEVRLNDELDTKGYTSLFLQRPDLRENGTJEDMFVYLGKND- 543  
QY 461 KLGQAMSHNVALLVREDPHGVCLESGDPT 492  
DB 544 -----ASKDYIGIGGCRMPADVCLQPGPRS 569

RESULT 14  
US-09-651-200-6  
Sequence 6, Application US/09651200  
Patent No. 6429303  
GENERAL INFORMATION:  
APPLICANT: Green et al  
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
TITLE OF INVENTION: Polypeptides Encoded Thereby  
FILE REFERENCE: 15966-562 (CURA-62)  
CURRENT APPLICATION NUMBER: US/09/651,200  
CURRENT FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152383  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/172909  
PRIOR FILING DATE: 1999-12-21  
PRIOR APPLICATION NUMBER: 60/183578  
PRIOR FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 534  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-651-200-6

Query Match 3.98; Score 101.5; DB 4; Length 534;  
Best Local Similarity 22.88; Pred. No. 0.12;  
Matches 97; Conservative 48; Mismatches 136; Indels 145; Gaps 23;

QY 183 GSSAAMTAVVALLHL-----LGVVDLSDPCKEKGCG-----SDLDVTHMI 225  
DB 86 GSAVANRTALFPDLAOGNASLRQRYRVAD---EGSFTCFVSIRDFGSAVSLQVAAP 142  
QY 226 AOTSCHLAQCK---VSGSFDVSCAVYSGQRYRFSPEVLSFAQVAVTGLPINEVITILK 282  
DB 143 SKPSMTLEPKDLRPGDTVITCSSY--QGY-----PEAEVFMQDG-QGVPLT--GNVTT 192  
QY 283 GKMDNKRTEFSPLPLMLFLGEPG-----SGSSSTP--SMGAVKKQW 324  
DB 193 SQMANEGELPDVHSIILRVVANGTYSCLVNRPVLAQDAHSVTITPQRSPTGAVEQVP 252  
QY 325 SDEPKAREMNQNSDANLELETKLNDLSLAKDMDVLYLVISCSVLTSEKVVLAH-TE 383  
DB 253 EDPVVALVG---TDATLRCSFSPBEPGSLAQ-----LNLWQ---LDTTKOLVHSFTE 299  
QY 384 PINE-----AIKELREAREAMLRIR-----ILMRMGEA----- 414  
DB 300 GROGSAVANRTALFPDLAOGNASLRQRYRVADDESSFTCFVSIRDFGSAVSLQVAAP 359  
QY 415 -----SVPIEPESQTLLDST-----MSAEVLLAG-----VEGAGG 446  
DB 360 YSKPSMTLEPKDLRPGDTVITCSSYRGYPEAEVEFMQDGQVPLTGNVTTSQMANEGL 419  
QY 447 FD-AIFAITLGDSTGLTQAMSHNVALLVR-----EDPHGVCLESGDPT-----TC 494  
DB 420 FDVHSAVRLVANGT-----YSCLVNRPVLAQDAHSVTITGQPMTFPEALM 468  
QY 495 ITSGVS 500  
DB 469 VFWGLS 474

RESULT 15  
US-09-651-200-24  
Sequence 24, Application US/09651200  
Patent No. 6429303  
GENERAL INFORMATION:  
APPLICANT: Green et al  
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
TITLE OF INVENTION: Polypeptides Encoded Thereby  
FILE REFERENCE: 15966-562 (CURA-62)  
CURRENT APPLICATION NUMBER: US/09/651,200  
CURRENT FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152383  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/172909  
PRIOR FILING DATE: 1999-12-21  
PRIOR APPLICATION NUMBER: 60/183578  
PRIOR FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 534  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Sequence  
OTHER INFORMATION: mz5020.protein from Figure 4.  
US-09-651-200-24

Query Match 3.98; Score 101.5; DB 4; Length 534;  
Best Local Similarity 22.88; Pred. No. 0.12;  
Matches 97; Conservative 48; Mismatches 136; Indels 145; Gaps 23;

QY 183 GSSAAMTAVVALLHL-----LGVVDLSDPCKEKGCG-----SDLDVTHMI 225  
DB 86 GSAVANRTALFPDLAOGNASLRQRYRVAD---EGSFTCFVSIRDFGSAVSLQVAAP 142



```

0Y 226 AOTSICLAQGG---VGSFSDYSCAVYGSQRRVRRSEPVLSFAQVAVYGLPLANEVICTILK 282
Db 143 SKPSTLTLPNNDLRRGDVTWITTCSSY--QGT---PEAEVFWODG--QGVLT---GNVVT 1922
0Y 283 GKMDNRKRFESLPLMLNLFLEGP-----SGGSSPP--SMGAVYKKMQM 324
Db 193 SOMANEQGLFVHSLRVGLNANTYSCLVARNPYLOQDAHSSVTIIPQRSPTCAVEYQAP 252
0Y 325 SDPEKARENNONLS DANLEETKLNLSIKLAKDHMDVYLLRVIKSCSVLTSEKVVLAH-TE 383
Db 253 EDPVVALVG---TDATLRCSFSEPEGFSLAQ-----LMLIMQ--LFTDKQLVHSFE 209
0Y 384 PINE-----AIIKELLEAREMLAIR-----ILMROMGAA----- 414
Db 300 GRDQGSAYANFTALFPELLAGNNSLRLOVRVRADEGSCFTCFYSIRPFGSAAYSLQVAP 359
0Y 415 ---SVPIPEESOTQLDST-----MSAEGVYLAG-----VPGAG 446
Db 360 YSKPSMTLEPRKKDLRRGDTVITTCSSYRGYPEAEVFWODGQGVPLTGNVITTSQMANEQGL 419
0Y 447 FD--AIFAITLGDGCTKLQAMSSHNVALIVR-----EDPHGVCSGSDPT-----TC 494
Db 420 FDIHVSLRVGLANGT-----YSCLVARNPYLOQDAHGSVITIGQPMTFPEPALM 468
0Y 495 ITSGVS 500
Db 469 VTVGLS 474

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 12:57:10 ; Search time 19 Seconds  
(without alignments)  
2129.766 Million cell updates/sec

Title: US-09-988-863a-2

Sequence: 1 MAVVASAPGKVLMTGTYLVL.....ESGDPRTTCITGVSIIHL 505

Scoring table: -BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2581	100.0	505	10	US-09-988-863a-2
2	575	22.3	179	10	US-09-909-745-22
3	453.5	17.6	451	10	US-09-909-745-24
4	216	8.4	67	10	US-09-909-745-20
5	204	7.9	54	10	US-09-909-745-16
6	204	7.9	54	10	US-09-909-745-18
7	178.5	6.9	368	10	US-09-815-242-10656
8	135.5	5.2	358	10	US-09-815-242-12866
9	133.5	5.2	345	10	US-09-815-242-5854
10	126	4.9	386	10	US-09-909-745-26
11	123	4.8	378	10	US-09-909-745-23
12	115	4.5	432	9	US-09-925-388-8
13	108.5	4.2	1477	9	US-10-092-880-4
14	101.5	3.9	534	9	US-10-077-023-7
15	101.5	3.9	534	10	US-09-875-338-7
16	101	3.9	1025	8	US-08-910-386a-7
17	100	3.9	335	10	US-09-815-242-13670
18	100	3.9	698	9	US-10-077-023-9
19	100	3.9	698	10	US-09-875-338-9

20	97	3.8	863	10	US-09-946-239-11	Sequence 11, Appl
21	96	3.7	657	10	US-09-815-242-13436	Sequence 13436, A
22	96	3.7	854	9	US-09-770-107-2	Sequence 2, Appl1
23	95.5	3.7	1599	9	US-10-092-880-9	Sequence 9, Appl1
24	95	3.7	492	9	US-09-738-626-5963	Sequence 5963, Ap
25	94	3.6	387	10	US-09-789-561-156	Sequence 156, App
26	93.5	3.6	806	12	US-10-003-405-2	Sequence 4, Appl1
27	92.5	3.6	514	9	US-09-738-626-5869	Sequence 5869, Ap
28	92.5	3.6	640	10	US-09-918-951-4	Sequence 4, Appl1
29	92.5	3.6	948	10	US-10-267-311-21	Sequence 21, Appl
30	92	3.6	385	10	US-09-853-918-37	Sequence 37, Appl
31	92	3.6	532	9	US-09-966-614-2	Sequence 2, Appl1
32	91.5	3.5	1176	9	US-09-918-508-2	Sequence 2, Appl1
33	91	3.5	679	10	US-09-815-242-5407	Sequence 5407, Ap
34	91	3.5	714	9	US-10-098-807-2	Sequence 2, Appl1
35	91	3.5	730	10	US-09-815-242-112489	Sequence 12489, A
36	91	3.5	1386	10	US-09-866-582-38	Sequence 38, Appl
37	90.5	3.5	351	9	US-10-078-770-138	Sequence 158, App
38	90.5	3.5	1165	10	US-09-815-242-10744	Sequence 10744, A
39	90.5	3.5	7968	9	US-10-077-130-5	Sequence 5, Appl1
40	89.5	3.5	794	10	US-09-815-242-11829	Sequence 11829, A
41	89	3.4	1173	9	US-10-135-322-19	Sequence 19, Appl
42	89	3.4	3313	10	US-09-737-149-29	Sequence 29, Appl
43	88.5	3.4	351	10	US-09-897-214-8	Sequence 8, Appl1
44	88.5	3.4	2344	10	US-09-815-242-12713	Sequence 12713, A
45	88	3.4	633	10	US-09-815-242-13318	Sequence 13318, A

#### ALIGNMENTS

RESULT 1  
US-09-988-863a-2  
; Sequence 2, Application US/09988863a  
; Patent No. US20020123427A1  
; GENERAL INFORMATION:  
; APPLICANT: Bayer AG  
; TITLE OF INVENTION: Plant phosphomevalonate kinases  
; FILE REFERENCE: Ie A 35 018  
; CURRENT APPLICATION NUMBER: US/09/988, 863a  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 505  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-988-863a-2

Query Match	100.0%	Score 2581	DB 10	Length 505
Best local similarity	100.0%	Pred. No. 9.5e-217		
Matches 505	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MAVVASAPGKVLMTGTYLLEKPNAGLVSTNARFAIYVINEEYKPEESMANKMTDVKL	60	
DB	1	MAVVASAPGKVLMTGTYLLEKPNAGLVSTNARFAIYVINEEYKPEESMANKMTDVKL	60	
QY	61	TSPOLRESMYKSLNLTLOSASASDSRNPFEHAIOVAIAAHLATERDKESLHKLL	130	
DB	61	TSPOLRESMYKSLNLTLOSASASDSRNPFEHAIOVAIAAHLATERDKESLHKLL	120	
QY	121	QGLDITLGSNDPYSYRNQIESAGLPTPESTLTLPFASTFTEMAESNGANSKPEYAKT	180	
DB	121	QGLDITLGSNDPYSYRNQIESAGLPTPESTLTLPFASTFTEMAESNGANSKPEYAKT	180	
QY	181	GLSSAAMTAVVAALHLYGVVDLSDPCKEGRGSDLDVIMIAQTSICLAQGVGSG	240	
DB	181	GLSSAAMTAVVAALHLYGVVDLSDPCKEGRGSDLDVIMIAQTSICLAQGVGSG	240	
QY	241	FDVSCAVYGSORVYRSPPEVLSPFAOVAVTGLPLNEVIGTILKGKMDKRTFEFSLPLML	300	
DB	241	FDVSCAVYGSORVYRSPPEVLSPFAOVAVTGLPLNEVIGTILKGKMDKRTFEFSLPLML	300	

QY	301	FLGEPGSGSGSTPSMWGAVYKKWQMSDEPKARENMÖNSDANLELETYLANDSLKLANDHMD	360
Db	301	FLGEPGSGSGSTPSMWGAVYKKWQMSDEPKARENMÖNSDANLELETYLANDSLKLANDHMD	360
QY	361	YLRYVYKCSVLTSEKMWLHATEDINEAIIKELLEAREAMLRIIRILMRONGEAASVIEP	420
Db	361	YLRYVYKCSVLTSEKMWLHATEDINEAIIKELLEAREAMLRIIRILMRONGEAASVIEP	420
QY	421	ESQÖQLDSTYMSAGSVLLAAGYPGAGGDFDAIFATILGDSGTRKLTQWMSHHVALLVREDP	480
Db	421	ESQÖQLDSTYMSAGSVLLAAGYPGAGGDFDAIFATILGDSGTRKLTQWMSHHVALLVREDP	480
QY	481	HGVCLESQDPRTTCITSGVSSIHLE	505
Db	481	HGVCLESQDPRTTCITSGVSSIHLE	505

RESULT 2  
 US-09-909-745-22  
 : Sequence 22, Application US/09909745  
 : Patent No. US20020119546A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Falco, S. Carl  
 : TITLE OF INVENTION: Squalene Synthesis Enzymes  
 : FILE REFERENCE: BB1112 US CIP  
 : CURRENT APPLICATION NUMBER: US/09/909,745  
 : CURRENT FILING DATE: 2001-07-20  
 : PRIOR APPLICATION NUMBER: 60/107,241  
 : PRIOR FILING DATE: 1998-11-05  
 : PRIOR APPLICATION NUMBER: 60/107,241  
 : PRIOR FILING DATE: 1998-11-05  
 : NUMBER OF SEQ ID NOS: 26  
 : SOFTWARE: Microsoft Office 97  
 : SEQ ID NO 22  
 : LENGTH: 179  
 : TYPE: PRT  
 : ORGANISM: *Oryza sativa*  
 US-09-909-745-22

	Query Match	Similarity	Best Local	Matches	Conservative	Mismatches	Indels	Gaps
	22.3%	Score 575;	DB 10;	length 179;				
	62.0%	Pred. No. 1,8e-42;						
	111;	Conservative	26;	Mismatches	42;	Indels	0;	Gaps 0;
Qy	308	GGSSPPSWGAYKKWQMSDPEKARBNWONI,SDANLELETKLND,SKLADHDHY,LRVIK	367					
Db	1	GGSSPPSWGAYKKWQMSDPEKARBNWONI,SDANLELETKLND,SKLADHDHY,LRVIK	60					
Qy	368	SCSVITSEKWLHATPEIPNEALIKELLEARREMLKRIIRLROMGEAASVPIEBSOTOLL	427					
Db	61	SCSVITSEKWLHATPEIPNEALIKELLEARREMLKRIIRLROMGEAASVPIEBSOTOLL	120					
Qy	428	DSTMAGEVLLAGVGAGGFDIAFATLIGDSTKTLQAASSHNVALLVLRDPHGCLE	486					
Db	121	DATMMNEGVLLAGVGAGGFDIAFATLIGDSTKTLQAASSHNVALLVLRDPHGCLE	179					

RESULT 3  
 US-09-909-745-24  
 Sequence 24, Application US/09909745  
 Patent No. US20020119546a1  
 GENERAL INFORMATION:  
 APPLICANT: Falco, S. Carl  
 APPLICANT: Famodu, Omolayo O.  
 TITLE OF INVENTION: Squelene Synthesis Enzymes  
 FILE REFERENCE: BB112 US CIP  
 CURRENT APPLICATION NUMBER: US/09/909,745  
 CURRENT FILING DATE: 2001-07-20  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 NUMBER OF SEQ ID NOS: 26

```

; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 451
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
; US-09-909-745-24

```

Query Match	17.6%;	Score 453.5;	DB 10;	Length 451;
Best Local Similarity	30.1%;	Pred. No. 2.9e-31;		
Matches 155;	Conservative 72;	Mismatches 185;	Indels 103;	Gaps 18

[illegible]

```

RESULT
US-09-909-745-20
: Sequence 20, Application US/09909745
: Patent No. US20020119546A1
: GENERAL INFORMATION:
: APPLICANT: Falco, S. Carl
: TITLE OF INVENTION: Squelene Synthesis Enzymes
: FILE REFERENCE: BB112 US CIP
: CURRENT APPLICATION NUMBER: US/09/909, 745
: PRIOR FILING DATE: 2001-07-20
: PRIOR APPLICATION NUMBER: 60/107, 241
: PRIOR FILING DATE: 1998-11-05
: PRIOR APPLICATION NUMBER: 60/107, 241
: PRIOR FILING DATE: 1998-11-05
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Microsoft Office 97
: SEQ ID NO: 20
: LENGTH: 67
: TYPE: prt
: ORGANISM: Oryza sativa
US-09-909-745-20

```

Query Match	8.4%;	Score 216;	DB 10;	Length 677
Best Local Similarity	60.0%;	Pred. No. 8.5e-12;		

Matches 39; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

OY 309 GSSPPSWGAVKRWMSDEPKARENMONTSDANLETKLNDLSKIKRKHDDYLRVTKS 368  
Db 1 GSSPPSWGAVKRWMSDEPKARENMONTSDANLETKLNDLSKIKRKHDDYLRVTKS 60

OY 369 CSVLT 373  
Db 61 CSRLT 65

RESULT 5  
US-09-909-745-16

; Sequence 16, Application US/09909745  
; Patent No. US20020119546A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Famodu, Omolayo O.  
; TITLE OF INVENTION: Squalene Synthesis Enzymes  
; FILE REFERENCE: BB1112 US CIP  
; CURRENT APPLICATION NUMBER: US/09/909,745  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 54  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-909-745-16

Query Match 7.9%; Score 204; DB 10; Length 54;  
Best Local Similarity 74.1%; Pred. No. 6.8e-11;  
Matches 40; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 397 REAMLRIRILKRMGEAASVPIEPESOTQLDSTMSAGVLLAGVPGAGGPDAT 450  
Db 1 RDACTLRILHREMGIAAGVPIEPESOTRLDATTMMEGVLLAGVPGAGGPDAT 54

RESULT 6  
US-09-909-745-18

; Sequence 18, Application US/09909745  
; Patent No. US20020119546A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Famodu, Omolayo O.  
; TITLE OF INVENTION: Squalene Synthesis Enzymes  
; FILE REFERENCE: BB1112 US CIP  
; CURRENT APPLICATION NUMBER: US/09/909,745  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 60/107,241  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 54  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-909-745-18

Query Match 7.9%; Score 204; DB 10; Length 54;  
Best Local Similarity 74.1%; Pred. No. 6.8e-11;  
Matches 40; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 337 REAMLRIRILKRMGEAASVPIEPESOTQLDSTMSAGVLLAGVPGAGGPDAT 450  
Db 1 RDACTLRILHREMGIAAGVPIEPESOTRLDATTMMEGVLLAGVPGAGGPDAT 54

RESULT 7  
US-09-815-242-10656

; Sequence 10656, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10656  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10656

Query Match 6.9%; Score 178.5; DB 10; Length 368;  
Best Local Similarity 22.3%; Pred. No. 2.1e-07;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;

OY 3 VVASAPGKVLMTGTYLEKPNAGLVSTNARFAIVPINEVPEPMARKMTIVTKLS 62  
Db 2 IEVTPGKFLTAGYAVVPEGPALIVADQFVTVVEETDEGSIQSAOYSSLPRT 60  
OY 63 POLRESMYKLSLNLTLQSVASDSRNPFEVHAIOVAIAAHATKDKESLHKLLOG 122  
Db 61 ---RNGELVDIR-----ENPF-----HYVLAHL-TEKTAOQNKEL--- 96  
OY 123 LDITLLGNDPFSYRNQIESAGLPLTPSLGLTAPASITFNAASNGANSKPEYAKTGL 182  
Db 97 -----SFYHLK-----VTSLSDSNG-----RKYGL 117  
OY 183 GSSAMTTAAVAAL--LHYLGAVDLSDECKEKEGCSLDVYHMAQTSKLAQKVS 240  
Db 118 GSSGAVTVGTAKALNIFDGL-----ENEELFKLSALAHAVQGN-GSC 161  
OY 241 FDVSCAVVGSORVREPEVLSFAQAVATGLPLNEVIGITLKGMKNRT-EFSLPLMN 299  
Db 162 GDIAASCTGG--WIAFTFDHDMVQKTY-----TETLIDLAMPPELMIPPLAKPKQLR 215  
OY 300 LFLGPPGGSGSTPSMGAVKRWMSDEPKARENMONTSDANLETKLNDLSKIKRKH 359  
Db 216 LLIGWTGS-PASTSDLVDRV--HQSKEKQAAVQQLKSKSLCYETMINGN----- 264  
OY 360 DYTLAVTSCSVLISEKVLNATEPINEATIKELLEAREAMLRIRILKRMGEAASV 419  
Db 265 -----TGRIS-VIOKQITKRNQ-----LTAELSSLTGVVIE 294

OY 420 PESOTQLDSTMSABEVLGAVPGAGFPAIFAIITLGDSC-TKLTQAMSHNVLAT 474  
DB 295 TEALKNLCIDIAESYTGA--AKSSGAGGDCGIYIFROKSGILPLMTAMEKDGITPL 348

## RESULT 8

US-09-815-242-12866  
Sequence 12866, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIORITY FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: 60/191,078  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY APPLICATION NUMBER: 60/206,848  
PRIORITY FILING DATE: 2000-05-23  
PRIORITY APPLICATION NUMBER: 60/207,727  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: 60/242,578  
PRIORITY FILING DATE: 2000-10-23  
PRIORITY APPLICATION NUMBER: 60/253,625  
PRIORITY FILING DATE: 2000-11-27  
PRIORITY APPLICATION NUMBER: 60/257,931  
PRIORITY FILING DATE: 2000-12-22  
PRIORITY APPLICATION NUMBER: 60/269,308  
PRIORITY FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for windows Version 4.0  
SEQ ID NO 12866  
LENGTH: 358  
TYPE: PRF  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12866

Query Match 5.2%; Score 135.5; DB 10; Length 358;  
Best Local Similarity 22.6%; Pred. No. 0.0011;  
Matches 108; Conservative 58; Mismatches 178; Indels 133; Gaps 25;

OY 3 VVASAPGKVLMTGGVLYLEKPNAGVLSTNARFYAIKPIINEEYKPSMAWKMTDVKLT 62  
DB 2 IQVAKGCKLTIVAGEYATE--PGYKSVLIALDRF--VTATIEE-----TD----- 42  
OY 63 POLSRESMYLSTLNHLTLOSASDSRNPVEHAIOYAIANAHLATEPKESLHKLLD 122  
DB 43 -----QIKGTHSKALHNPVTFSRD--EDSI--VISDPHAKQ-----TD----- 77  
OY 123 LDTITLGSNDYFYSRNOIESA--GLPLTPESLGTLPASFASITFNAESNGANSKPEVAKT 180  
DB 78 LNVVVAIEIFEQYAKSCDIAMKHFHLLTDS-----NLDSNG-----HKY 118  
OY 181 GLGSSAAMTTAAVVAALHLYGVVDLSDPCKEKGCSDDLVIHMLAOTSCHLAQKVGSG 240  
DB 119 GLGSSAVALVSVIKVINEFDM-----KLSNL--YIKYLAIVAMKLOSLSSCG 165  
OY 241 FDSVCAY--GSQRYVRPSPVLSFAQAVYGLPLNEVIGTILKGKMDNKRTF--FSLPPLM 298  
DB 166 -DIAVAVSGWLAISTFDHEWVKH-QIEDT-----TVEVLLIKMNGGLHIEPLQAPENM 217  
OY 299 NLELGPSSGSGSTPMSVAGVKKWMSDPEKARENMONLSDANLETKLNDLSKLAKDH 358  
DB 218 EVLIGMTGSPASS--PHFVSEVKRLK--SDP-----SF 246

OY 359 WDYLRYVISCVSLTSEKVLHATEPINEAITELLEAREAMRIRILMROGEAASVPI 418  
DB 247 YCDFLEDSHRC-----VEK-LIHAFTNNIKGVQKVRON-----RTIIORMKREATVDI 295  
OY 419 EPESQQLDSTMSABEVLGAVPGAGFPAIFAIITLGD--SGTKLQAMSHNVLAT 474  
DB 296 ETEKLYLCIDIAEKYHGA--SKTSGAGGDCGITTINKVDYDKERTIDEMTKHGKIKPL 350

## RESULT 9

US-09-815-242-5854  
Sequence 5854, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIORITY FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: 60/191,078  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY APPLICATION NUMBER: 60/206,848  
PRIORITY FILING DATE: 2000-05-23  
PRIORITY APPLICATION NUMBER: 60/207,727  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: 60/242,578  
PRIORITY FILING DATE: 2000-10-23  
PRIORITY APPLICATION NUMBER: 60/253,625  
PRIORITY FILING DATE: 2000-11-27  
PRIORITY APPLICATION NUMBER: 60/257,931  
PRIORITY FILING DATE: 2000-12-22  
PRIORITY APPLICATION NUMBER: 60/269,308  
PRIORITY FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for windows Version 4.0  
SEQ ID NO 5854  
LENGTH: 345  
TYPE: PRF  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5854

Query Match 5.2%; Score 133.5; DB 10; Length 345;  
Best Local Similarity 22.8%; Pred. No. 0.0016;  
Matches 107; Conservative 57; Mismatches 173; Indels 133; Gaps 25;

OY 7 APKVLMTGGVLYLEKPNAGVLSTNARFYAIKPIINEEYKPSMAWKMTDVKLTSPQ 66  
DB 3 APKVLMTGGVLYLEKPNAGVLSTNARFYAIKPIINEEYKPSMAWKMTDVKLTSPQ 66  
OY 67 RESMYLSTLNHLTLOSASDSRNPVEHAIOYAIANAHLATEPKESLHKLLDGLDT 126  
DB 40 ---QYKGTIHSKALHNPVTFSRD--EDSI--VISDPHAKQ-----LNV 78  
OY 127 ILGSDNYFYSRNOIESA--GLPLTPESLGTLPASFASITFNAESNGANSKPEVAKTGLG 184  
DB 79 VTAIEIFEQYAKSCDIAMKHFHLLTDS-----NLDSNG-----HKYGLG 119  
OY 185 SAAMTTAAVVAALHLYGVVDLSDPCKEKGCSDDLVIHMLAOTSCHLAQKVGSGPDV 244  
DB 120 SAVALVSVIKVINEFDM-----KLSNL--YIKYLAIVAMKLOSLSSCG--DIA 165  
OY 245 CAVY--GSQRYVRPSPVLSFAQAVYGLPLNEVIGTILKGKMDNKRTF--FSLPPLANLFL 302  
DB 166 VSVSGWLAISTFDHEWVKH-QIEDT-----TVEVLLIKMNGGLHIEPLQAPENNEVL 218

QY 303 GEPGSGSSTPSMAYGAVKKQMSPDKARENMOMLSANLELETKLNDLSKLAKDHMDY 362  
 Db 219 GWTGSPASS-PHFVSEKRLK-SDP-----SFYDGF 247  
 QY 363 LRVKSCSVLTSEKVLHATEPINEAIKELLEAREAMLRIILRMONGEASVPIEPES 422  
 Db 248 LEDSHRC-----VER-LIHAFTNNIKGVQKVVRON-----RTIIRMDKEATVDIETEK 296  
 QY 423 QTOLDSTMSAEVLLAGVPGAGFDALFATLLD-SGTRKTQAMSSINV 471  
 Db 297 LKYLCLDAEKYHGA-SKTSAGGGGDCITILNKVDKREKITYDEWTKHGI 344

# RESULT 10 US-09-909-745-26

Sequence 26, Application US/09909745  
 Patent No. US20020119546a1  
 GENERAL INFORMATION:  
 APPLICANT: Falco, S. Carl  
 APPLICANT: Famodu, Omolayo O.  
 TITLE OF INVENTION: Squelene Synthesis Enzymes  
 FILE REFERENCE: BB112 US CIP  
 CURRENT APPLICATION NUMBER: US/09/909,745  
 PRIOR FILING DATE: 2001-07-20  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: Microsoft Office 97  
 SEQ ID NO 26  
 LENGTH: 386  
 TYPE: PRF  
 ORGANISM: Glycine max  
 US-09-909-745-26

Query Match 4.9%; Score 126; DB 10; Length 386;  
 Best Local Similarity 21.0%; Pred. No. 0.0084;  
 Matches 97; Conservative 73; Mismatches 155; Indels 136; Gaps 22;

QY 7 APGVNMTGCVLYLEKPNAGLVSTNARFYAIVKPINEEVPESNAKMTDVKLTSPOLS 66  
 Db 6 APGVNMTGCVLYLEKPNAGLVSTNARFYAIVKPINEEVPESNAKMTDVKLTSPOLS 66  
 QY 67 RESMYKLSLNHLTLO-SVASDSRNPFVEHAIQYAIAMHATEKDESKLKLGLD 125  
 Db 46 EDLSL-KLKLOTTALEFPSPITRIRAPPESTAOUSSTPNSCSVENAKAI--AAIVEELNI 102  
 QY 126 TILGSDNDFSYRNOIESAGLPLTPES-LGTLPASITFNAESNGANSKPEPAK----- 179  
 Db 103 -----PEAKGLASVSAPFLMTLSIOGFRPAIVVTSLEPL 139  
 QY 180 -TGIGSSAAMTAVVALLHLYGVVDSDPCKEG--KEGSDLDVIMIAOTSCHLAQK 236  
 Db 140 GSGGSSSFCVLAALALAYTDSVSL-DKHQGMISFGEKDLPLYNMAFEKEXIIRK 198  
 QY 237 VGSFEDVSCAVYGSQRYRFPSEVLISFAQVAVT--GLPLNEVIGTILKKMKNKREF 292  
 Db 199 -PSIDITVSAIGN-----ITISFGSNLTHMKSSVPLKML-----TNTKV- 238  
 QY 293 SLPLMLNLFGEPSGSGSSTPSMAYGAVKKQMSPDKARENMOMLSANLELETKLNDLS 352  
 Db 239 -----GRNTKALVAGVGERMLRHPD-----IMAFVSADVDSINLELT 275  
 QY 353 KLAQDHMDVILRVKSCSVLTSEKVLHATEPINEAIKELLEAREAMLRIILRMONGE 412  
 Db 276 SILKSP-----TPDE--LSYTE--KEEKIEELMENNOGML-----QSMG- 310  
 QY 413 AASVPIEPESOTOLDSTMSAEVLLAGVPGAGFDALFAT 453  
 Db 311 VSHAETIETVLTIL-----KYKLASKLTGAGGGCVLT 344

# RESULT 11 US-09-909-745-23

Sequence 23, Application US/09909745  
 Patent No. US20020119546a1  
 GENERAL INFORMATION:  
 APPLICANT: Falco, S. Carl  
 APPLICANT: Famodu, Omolayo O.  
 TITLE OF INVENTION: Squelene Synthesis Enzymes  
 FILE REFERENCE: BB112 US CIP  
 CURRENT APPLICATION NUMBER: US/09/909,745  
 PRIOR FILING DATE: 2001-07-20  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 PRIOR APPLICATION NUMBER: 60/107,241  
 PRIOR FILING DATE: 1998-11-05  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: Microsoft Office 97  
 SEQ ID NO 23  
 LENGTH: 378  
 TYPE: PRF  
 ORGANISM: Arabidopsis thaliana  
 US-09-909-745-23

Query Match 4.8%; Score 123; DB 10; Length 378;  
 Best Local Similarity 20.9%; Pred. No. 0.015;  
 Matches 100; Conservative 76; Mismatches 151; Indels 150; Gaps 23;

QY 1 MAVVASAPGVNMTGCVLYLEKPNAGLVSTNARFYAIVKPINEEVPESNAKMTDVKLT 60  
 Db 1 MEVVARAPGVNMTGCVLYLEKPNAGLVSTNARFYAIVKPINEEVPESNAKMTDVKLT 60  
 QY 61 TSPQSRSMYKLSLNHLTLOSVASDSRN-PVEHAIQYAIAMHATEK--DKESK 117  
 Db 39 REPVLSAENNDRLTLO--LKQISLEFWSLARIEALIPYDSITLCKSTPAQSCSEITLK 94  
 QY 118 LLOGGLDITILGSDNDFSYRNOIESAGLPLTPESLGT-LAPF-----ASITFN 164  
 Db 95 -----STAVL-----VEQNLPRKEMMLSSGISTFMYLTRIIGENPATVIN 137  
 QY 165 AAEBSGANSKPEPAKGTGIGSSAAMTAVVALLHLYGVVDSDPCKEKEFGCD--LDV 221  
 Db 138 SELPYG-----SGGSSAALCVALLTAAL-----ASSISEKTRGNGWSLDETNEL 184  
 QY 222 IHMTAOTSCHLAQGVGSGFVSCAVYGSQRYRFPSEVLISFAQVAVGTLPLNEVIGTIL 281  
 Db 185 LNKAPFEKEXIIRK-PGIDNTVSAIGN-----MIFCSEITRLQSNPLRLI 234  
 QY 282 KGMKDNKRTFESLPLMLNLFGEPSGSGSSTPSMAYGAVKKQMSPDKARENMOMLSAN 341  
 Db 235 -----TNTRV-----GRNTKALVSGVQRAVRHPDAMKSVFNAVDSIS 272  
 QY 342 LELETKLNDLSKLAKDHMDVILRVKSCSVLTSEKVLHATEPINEAIKELLEAREAML 401  
 Db 273 KELAIIID--SKDSTV-----TEK-----EERIKELMENNOGML 305  
 QY 402 RIRILRMONGEASVPIEPESOTOLDSTMSAEVLLAGVPGAGFDALFATLLDSTG 460  
 Db 306 -----LSMGVSHSIEAVIITTVAKH--LVSKLIGAGGGCV--LTLLPTGT 348

# RESULT 12 US-09-925-388-8

Sequence 8, Application US/09925388  
 Publication No. US20030054523A1  
 GENERAL INFORMATION:  
 APPLICANT: HOSHINO, Tatsuo  
 APPLICANT: OJIMA, Kazuyuki  
 APPLICANT: SETOGUCHI, Yutaka  
 TITLE OF INVENTION: ISOPRENOID PRODUCTION  
 FILE REFERENCE: ISOPRENOID PRODUCTION  
 CURRENT APPLICATION NUMBER: US/09/925,388

; PRIOR APPLICATION NUMBER: 09/306,595  
 ; PRIOR FILING DATE: 1999-05-06  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 432  
 ; TYPE: PRT  
 ; ORGANISM: Phaffia rhodozyma  
 ; US-09-925-388-8

Query Match 4.5%; Score 115; DB 9; Length 432;

Best Local Similarity 21.6%; Pred. No. 0.091; Indels 120; Gaps 19;

Matches 103; Conservative 64; Mismatches 190; Indels 120; Gaps 19;

QY 3 VVASAPGVMTGGYLVLEKPNAGLVLTNARFAIYKPIINEEVKPESMAMKTDVLTLS 62  
 DB 4 ILVSAPEKVLFGFHNH-GHGVGTGLASVDRCLALLSP-----TATTTTS 48  
 QY 63 POLSRESMYKLSLNLTLQSVASDSRNPEVHAIOYAIAMHLATEKD-KESLHKLLQ 121  
 DB 49 SLSLSTNI-TISLTLNFTQSWPVS-----LPMSLAPDW--TEASIPESICPTLLA 97  
 QY 122 GLD-TITLGSND-----FYSYRQIESAGLPPLPESIGTLAPASTIFMAESNGA 171  
 DB 98 EIERIAGGGNGGEREVATMAFLYLVLKSKGKPEPEL-----TANSALPMGA 148  
 QY 172 NSRPEVAKTGLSSAAMTTAVALLHYLGVVDLSDPCKEKGKFGCSDDLYIHIAQTSHC 231  
 DB 149 -----GLSSAALSTSLALVFLHFSLPTTGRESTIPTADTEVIDKWAFLAEK 199  
 QY 232 LAGKVGSGFDVSCAVYGSRRYVRFSEPVLSFAVANTGLPLNEVICTILKGDKNKRT 291  
 DB 200 VTHGN-PSGIDNNAVSTRGG-----AAVEKRIEKGQ-E 230  
 QY 292 FSLPPLMNL-----FLGEPGSGSSTPSMGAVKMKQMSDEK-----ARENMON 336  
 DB 231 GMEAIKSFISIRFLITDSIGRDTSLVAGVANRLQEPVIVPLEAQADEAIRC 290  
 QY 337 LSDANLEETKLDLSKLAKDHDVYLVKIKSCSVLTSEKWLHATEPIEAIKELLEA 396  
 DB 291 LKDSERAMVIMIDRLQNLVSEN-----HAHLAALGVSHPSLEIIR-IGAD 335  
 QY 397 REMLRIRILRMGGEASVPIEPESOTOLDSTM-----SABEVLLAGPAGG-GF 447  
 DB 336 KPEELRKLTKGAGGCAVTLVPDDSTETLLQALMETLVGSSPAPYARVGGSGVGF 392

# RESULT 13

; Sequence 4, Application US/10092880  
 ; Patent No. US20020164354A1

GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J.  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE  
 ; FILE REFERENCE: HAEMOPHILUS

CURRENT APPLICATION NUMBER: US/10/092,880

CURRENT FILING DATE: 2002-03-08

PRIOR APPLICATION NUMBER: 09/155,614

PRIOR FILING DATE: 1998-09-30

PRIOR APPLICATION NUMBER: 08/617,697

PRIOR FILING DATE: 1996-04-01

PRIOR APPLICATION NUMBER: PCT/US97/04707

PRIOR FILING DATE: 1997-04-01

NUMBER OF SEQ ID NOS: 11

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 4

LENGTH: 1477

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-092-880-4

Query Match 4.2%; Score 108.5; DB 9; Length 1477;

Best Local Similarity 20.3%; Pred. No. 2.2; Indels 119; Gaps 21;

Matches 89; Conservative 68; Mismatches 162; Indels 119; Gaps 21;

QY 11 VLMTGGYLVLEKPN-----GLVLTNARFAIYKPIINEEVKPESMAMKTDVLTLS 56  
 DB 139 ILDSNGOVFLINRNGITIGDAIINTNGFTASTLIDLSNENIKARNFTFEDTKKALAEIV 198  
 QY 57 -----DYKLTSPOLSRSMYKLSLNLTLQSVASDSRNPEVHAIOYAI 102  
 DB 199 NHGLITVGRKDSVNLIGGKVKNEGVSIVNGSISLLAGKITISDIINP-----TIFYSIA 254  
 QY 103 AAMLAETKDESLHKLGLDITLIGSNDIFYRQDIESAGLPPLPESIGTLAPASTIT 162  
 DB 255 -----APENEAENVLGDIFPAKGNINVRRA-----TIRNQGKLSADSVSKDKSGNTV--LSAK 304  
 QY 163 FNAESNGA-NSRPEVAKT-----LGSSAAMTTAVAAALL-----HYLGVVLDSPCKE 211  
 DB 305 EGEAEIGGVISAQNOQAKGKMLITGDKYVLKIGAVIDLSGKGGERYLG-----GDERGE 360  
 QY 212 GFRGCSDDLYIHIAQTSHC-----SHCLAQKVGSGFDVSCAVYGSRRYVRFSEPVLSFAOV 266  
 DB 361 GRNG-----IQLAKTSLKSGSTINVSKEKGFAL--VMGDIALIDGNINAQSGDI 411  
 QY 267 AVTG-----LPLNEVICTILKGM-----DNKRTFESLPLMNLFLGPGSGSS 311  
 DB 412 AKTGGEVETSGHDLFIKDNALVDA--KEWLLDPDNVSIABDPLENNGTINDEPPTGTG 468  
 QY 312 TPSMGAVKMKQMSDEKARENQNLSDANLEETKLD--LSKLAKDHDV-----YL 363  
 DB 469 -----EASDPK-----NSELKTLTNTTNTISNYLKNAMTMMITASRKL 506  
 QY 364 RVIKSCSVLTSEKWLHA 381  
 DB 507 TVNSINIGNSHLLILHS 524

# RESULT 14

Sequence 7, Application US/10077023

Publication No. US20030031675A1

GENERAL INFORMATION:

APPLICANT: MIKESSELL, GLEN E.

APPLICANT: CHANG, HAN

APPLICANT: FINGER, JOSHUA N.

APPLICANT: YANG, GUOCHEN

APPLICANT: LU, PIN

APPLICANT: ZHOU, XIA-DI

APPLICANT: PEACH, ROBERT

TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR

FILE REFERENCE: IMMUNOMODULATION

CURRENT APPLICATION NUMBER: US/10/077,023

CURRENT FILING DATE: 2002-02-15

PRIOR APPLICATION NUMBER: 60/272,107

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: 60/209,811

PRIOR FILING DATE: 2000-06-06

NUMBER OF SEQ ID NOS: 138

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 7

LENGTH: 534

TYPE: PRT

ORGANISM: Homo sapiens

US-10-077-023-7

Query Match 3.9%; Score 101.5; DB 9; Length 534;

Best Local Similarity 22.8%; Pred. No. 1.9; Indels 145; Gaps 23;

Matches 97; Conservative 48; Mismatches 136; Indels 145; Gaps 23;

QY 183 GSSAAMTTAVAAALHY-----LGVVLDSPCKEKGKFGC-----SDLDVYHMI 225

DB 86 GSAVANRNTALFPDLLAAGNLSRLQRYRVAD-----ESFTCFVSIIRDFGSAANVSLQYAAAP 142



QY 226 AOTSCHLAOCK---VSGGFVSCAVYGSQRYVRFSPVLSPFAQVAVTGLPLNEVIGTILK 282  
DB 143 SKPSMTLEPKKDLRPGDTVTITCSSY--QGY-----PEAEVFMQDG--QGVPLT---GNVTT 192  
QY 283 GKDNKRTESLPLPLNLFGEPCG-----SGGSSTP--SMGCAVKKWQM 324  
DB 193 SOMANEGGLDFVHSILRVVILGANGTYSCLVRNPVLAQDAHSSVTITPQRSPTGAVEQVP 252  
QY 325 SDEKARENQONLSDANLEETKLNDSKLAKDHWDYILRVIKSCSVLTSEKWLHA-TE 383  
DB 253 EDPVVALVG---TDATLRCSFSPEPGFSLAQ-----LNLIMO---LITDKLVHSFTE 299  
QY 384 PINE-----AIKELLEAREAMLIR-----ILMROMGEAA----- 414  
DB 300 GRDGSAYANRTALFPDLAOGNASLRLQVRVADEGSFTCFVSIROFGSAVSLQYAAP 359  
QY 415 ----SVPIEPSQTQLDST-----MSAEVLLAG-----VPGAGG 446  
DB 360 YSKPSMTLEPKKDLRPGDTVTITCSSYRGYPEAEVFMQDGCVPLTGNVTTSCMANEGL 419  
QY 447 FD--AIFAITLGDGSKTLTQAMSSHNVLALIVR-----EDPHGVCLESGDPRT-----TC 494  
DB 420 EDVHSVLRVILGANGT-----YSCLVNRPVLAQDAHSSVTITGQPMTPPPEALW 468  
QY 495 ITSGVS 500  
DB 469 VTGGLS 474

RESULT 15  
US-09-875-338-7  
; Sequence 7, Application US/09875338  
; Patent No. US20020095024A1  
; GENERAL INFORMATION:  
; APPLICANT: MIKESELL, GLEN E.  
; APPLICANT: CHANG, HAN  
; APPLICANT: FINGER, JOSHUA N.  
; APPLICANT: YANG, GUCHEN  
; APPLICANT: LU, PIN  
; APPLICANT: ZHOU, XIA-DI  
; APPLICANT: PEACH, ROBERT  
; TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR  
; FILE REFERENCE: 3053-4071US2  
; CURRENT APPLICATION NUMBER: US/09/875,338  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: 60/272,107  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/209,811  
; PRIOR FILING DATE: 2000-06-06  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-875-338-7

Query Match 3.9%; Score 101.5; DB 10; Length 534;  
Best Local Similarity 22.8%; Pred. No. 1.9;  
Matches 97; Conservative 48; Mismatches 136; Indels 145; Gaps 23;

QY 183 GSSAAMTTAVVALLHY-----LGVDLSDPKCKEKGFC-----SDLDVIHMI 225  
DB 86 GSAVYANRTALFPDLAOGNASLRLQVRVAD---EGSFTCFVSIROFGSAVSLQYAAPY 142  
QY 226 AOTSCHLAOCK---VSGGFVSCAVYGSQRYVRFSPVLSPFAQVAVTGLPLNEVIGTILK 282  
DB 143 SKPSMTLEPKKDLRPGDTVTITCSSY--QGY-----PEAEVFMQDG--QGVPLT---GNVTT 192  
QY 283 GKDNKRTESLPLPLNLFGEPCG-----SGGSSTP--SMGCAVKKWQM 324  
DB 193 SOMANEGGLDFVHSILRVVILGANGTYSCLVRNPVLAQDAHSSVTITPQRSPTGAVEQVP 252

QY 325 SDEKARENQONLSDANLEETKLNDSKLAKDHWDYILRVIKSCSVLTSEKWLHA-TE 383  
DB 253 EDPVVALVG---TDATLRCSFSPEPGFSLAQ-----LNLIMO---LITDKLVHSFTE 299  
QY 384 PINE-----AIKELLEAREAMLIR-----ILMROMGEAA----- 414  
DB 300 GRDGSAYANRTALFPDLAOGNASLRLQVRVADEGSFTCFVSIROFGSAVSLQYAAP 359  
QY 415 ----SVPIEPSQTQLDST-----MSAEVLLAG-----VPGAGG 446  
DB 360 YSKPSMTLEPKKDLRPGDTVTITCSSYRGYPEAEVFMQDGCVPLTGNVTTSCMANEGL 419  
QY 447 FD--AIFAITLGDGSKTLTQAMSSHNVLALIVR-----EDPHGVCLESGDPRT-----TC 494  
DB 420 EDVHSVLRVILGANGT-----YSCLVNRPVLAQDAHSSVTITGQPMTPPPEALW 468  
QY 495 ITSGVS 500  
DB 469 VTGGLS 474

Search completed: April 26, 2003, 13:02:22  
Job time : 22 secs

THIS PAGE BLANK (USPTO)



QY 121 QGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAKT 180  
 DB 121 QGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAKT 180  
 QY 181 GIGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 240  
 DB 181 GIGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 240  
 QY 241 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 300  
 DB 241 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 300  
 QY 301 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 360  
 DB 301 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 360  
 QY 361 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 420  
 DB 361 FVSCAVNGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 420  
 QY 421 ESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVREDP 480  
 DB 421 ESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVREDP 480  
 QY 481 HGVCLESDDPRTCTITSGVSIHLE 505  
 DB 481 HGVCLESDDPRTCTITSGVSIHLE 505

## RESULT 2

PCT-US02-24048A-12  
 ; Sequence 12, Application PC/TUS0224048A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: E.I. du Pont de Nemours & Company  
 ; TITLE OF INVENTION: Genes Involved in the Biosynthesis of Isopentenyl Diphosphate in  
 ; FILE REFERENCE: CL1792 PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US02/24048A  
 ; PRIOR FILING DATE: 2002-09-10  
 ; PRIOR APPLICATION NUMBER: 60/307,637  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 12  
 ; LENGTH: 503  
 ; TYPE: PRT  
 ; ORGANISM: Hevea brasiliensis  
 ; PCT-US02-24048A-12

Query Match 74.8%; Score 1930.5; DB 1; Length 503;  
 Best Local Similarity 74.1%; Pred. No. 5.9e-182;  
 Matches 374; Conservative 61; Mismatches 67; Indels 3; Gaps 2;

QY 1 MAVVASAPKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEYKPPSMAAMKTDVYL 60  
 DB 1 MAVVASAPKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEYKPPSMAAMKTDVYL 60  
 QY 61 TSPOLRESMYKLSLNHLTLOSASDSRNPFEYEAIOYAAAHATEKDKES-LHKLL 119  
 DB 61 TSPOLRESMYKLSLNHLTLOSASDSRNPFEYEAIOYAAAHATEKDKES-LHKLL 119  
 QY 120 LOGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAK 179  
 DB 120 LOGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAK 179  
 QY 180 TELGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 239  
 DB 180 TELGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 239  
 QY 240 GFDVSAVYGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 299  
 DB 240 GFDVSAVYGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 299

QY 300 LFLGPEGGSSSTPSPVAGVKKWQMSDPEKARENMONLSDANLETKLNDLSKLAKDH 359  
 DB 299 LFLGPEGGSSSTPSPVAGVKKWQMSDPEKARENMONLSDANLETKLNDLSKLAKDH 358  
 QY 360 DYLRYKSCSVLTSEKWLHATEPINEALIKELLEAREAMLRIRILMOMGEAASVPIE 419  
 DB 359 DAKCVYDSCSTYKSEKWLHATEPINEALIKELLEAREAMLRIRILMOMGEAASVPIE 418  
 QY 420 PESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVRED 479  
 DB 419 PESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVRED 478  
 QY 480 HGVCLESDDPRTCTITSGVSIHLE 504  
 DB 479 PNGVLLSDDPRTCTITSGVSIHLE 503

## RESULT 3

US-10-036-959B-12  
 ; Sequence 12, Application US/10036959B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: E.I. du Pont de Nemours & Company  
 ; APPLICANT: Hallahan, David L.  
 ; TITLE OF INVENTION: Genes Involved in the Biosynthesis of Isopentenyl Diphosphate  
 ; FILE REFERENCE: CL-1792  
 ; CURRENT APPLICATION NUMBER: US-10/036,959B  
 ; PRIOR FILING DATE: 2002-05-10  
 ; PRIOR APPLICATION NUMBER: 60/307,637  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 12  
 ; LENGTH: 503  
 ; TYPE: PRT  
 ; ORGANISM: Hevea brasiliensis  
 ; US-10-036-959B-12

Query Match 74.8%; Score 1930.5; DB 24; Length 503;  
 Best Local Similarity 74.1%; Pred. No. 5.9e-182;  
 Matches 374; Conservative 61; Mismatches 67; Indels 3; Gaps 2;

QY 1 MAVVASAPKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEYKPPSMAAMKTDVYL 60  
 DB 1 MAVVASAPKVLMTGGYLVLEKPNAGLVSTNARFYAIKPIINEYKPPSMAAMKTDVYL 60  
 QY 61 TSPOLRESMYKLSLNHLTLOSASDSRNPFEYEAIOYAAAHATEKDKES-LHKLL 119  
 DB 61 TSPOLRESMYKLSLNHLTLOSASDSRNPFEYEAIOYAAAHATEKDKES-LHKLL 119  
 QY 120 LOGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAK 179  
 DB 120 LOGDLITIIISNDYFYRYNOIESAGLPTEPESLGTAPFASITFNAESNGANSKREPVAK 179  
 QY 180 TELGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 239  
 DB 180 TELGSSAAMTTAVAAALHLHLYGVNDLSDPCKEKGFGCSDDIVHIMIAOTSHCLAQGVSG 238  
 QY 240 GFDVSAVYGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 299  
 DB 239 GFDVSAVYGSORRYRSPSEVLSPFAQVAVTGLPNEVIGTILKGMKNKRTESLPLMLN 298  
 QY 300 LFLGPEGGSSSTPSPVAGVKKWQMSDPEKARENMONLSDANLETKLNDLSKLAKDH 359  
 DB 299 LFLGPEGGSSSTPSPVAGVKKWQMSDPEKARENMONLSDANLETKLNDLSKLAKDH 358  
 QY 360 DYLRYKSCSVLTSEKWLHATEPINEALIKELLEAREAMLRIRILMOMGEAASVPIE 419  
 DB 359 DAYKCVYDSCSTYKSEKWLHATEPINEALIKELLEAREAMLRIRILMOMGEAASVPIE 418  
 QY 420 PESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVRED 479  
 DB 419 PESOTOLDSTMSAEGVLLAGVPAGGFDAIFAITLGDSTKLTQAMSSHNVALLVRED 478

Db 419 PESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 QY 480 PHGVLESQDPRTCITSGVSSIHL 504  
 Db 479 PNGVLESQDPRTCITTAFAVAHI 503

# RESULT 4 US-10-219-999-40847

; Sequence 40847, Application US/10219999  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Edgerton, Michael D  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Liu, Jindong  
 ; APPLICANT: Stein, Joshua  
 ; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
 ; FILE REFERENCE: 38-10(52726)C  
 ; CURRENT APPLICATION NUMBER: US/10/219,999  
 ; CURRENT FILING DATE: 2002-08-15  
 ; PRIOR APPLICATION NUMBER: US 60/324,109  
 ; PRIOR FILING DATE: 2001-09-21  
 ; PRIOR APPLICATION NUMBER: US 60/312,544  
 ; PRIOR FILING DATE: 2001-08-15  
 ; NUMBER OF SEQ ID NOS: 63520  
 ; SEQ ID NO 40847  
 ; LENGTH: 511  
 ; TYPE: PR1  
 ; ORGANISM: Zea mays  
 ; US-10-219-999-40847

Query Match 61.2%; Score 1579; DB 26; Length 511;  
 Best local Similarity 62.6%; Pred. No. 5.7e-147;  
 Matches 321; Conservative 69; Mismatches 111; Indels 12; Gaps 5;

QY 1 MAVVASAPGKVLMTGTYLLEKPNAGLVLTSTNARFYAIRPINEVPEESAMKMTDVKL 60  
 Db 1 MEVVASAPGKVLMTGTYLLEKPNAGLVLTSTNARFYAIRPINEVPEESAMKMTDVKV 60  
 QY 61 TSPQLSRESMYKLSLNHLTQSVASDSRNPFEVHAIOYAIANAHLATEKDKES---LHK 117  
 Db 61 TSPQLSRESMYKLSLNHLTQSVASDSRNPFEVHAIOYAIANAHLATEKDKES---LHK 117  
 QY 118 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 Db 118 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 QY 120 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 Db 120 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 QY 174 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 Db 174 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 QY 180 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 Db 180 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 QY 232 LAQKVGSGDFVSCAVYGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 291  
 Db 232 LAQKVGSGDFVSCAVYGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 291  
 QY 240 IAGKIGSGDFVSAAYVGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 297  
 Db 240 IAGKIGSGDFVSAAYVGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 297  
 QY 292 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 Db 292 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 QY 298 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 Db 298 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 QY 352 SKLAKDHMDVLYRIKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILMONG 411  
 Db 352 SKLAKDHMDVLYRIKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILMONG 411  
 QY 358 KGLSENHREAVESVYLSCHLTYEKTWEVATNOHQLIKSLAARDACLETIRLHMREMG 417  
 Db 358 KGLSENHREAVESVYLSCHLTYEKTWEVATNOHQLIKSLAARDACLETIRLHMREMG 417  
 QY 412 EASVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 Db 412 EASVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 QY 418 IAGVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 Db 418 IAGVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 QY 472 LALLVREDPHVCLESQDPRTCITSGVSSIHL 504  
 Db 472 LALLVREDPHVCLESQDPRTCITSGVSSIHL 504  
 QY 478 LPLLVREDPHVCLESQDPRTCITSGVSSIHL 510  
 Db 478 LPLLVREDPHVCLESQDPRTCITSGVSSIHL 510

# RESULT 5 US-60-324-109-16899

; Sequence 16899, Application US/603241095  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Edgerton, Michael D  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Liu, Jindong  
 ; APPLICANT: Stein, Joshua  
 ; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
 ; FILE REFERENCE: 38-10(52726)B  
 ; CURRENT APPLICATION NUMBER: US/60/324,109  
 ; CURRENT FILING DATE: 2001-09-21  
 ; NUMBER OF SEQ ID NOS: 33196  
 ; SEQ ID NO 16899  
 ; LENGTH: 511  
 ; TYPE: PR1  
 ; ORGANISM: Zea mays  
 ; US-60-324-109-16899

Query Match 61.2%; Score 1579; DB 27; Length 511;  
 Best local Similarity 62.6%; Pred. No. 5.7e-147;  
 Matches 321; Conservative 69; Mismatches 111; Indels 12; Gaps 5;

QY 1 MAVVASAPGKVLMTGTYLLEKPNAGLVLTSTNARFYAIRPINEVPEESAMKMTDVKL 60  
 Db 1 MEVVASAPGKVLMTGTYLLEKPNAGLVLTSTNARFYAIRPINEVPEESAMKMTDVKV 60  
 QY 61 TSPQLSRESMYKLSLNHLTQSVASDSRNPFEVHAIOYAIANAHLATEKDKES---LHK 117  
 Db 61 TSPQLSRESMYKLSLNHLTQSVASDSRNPFEVHAIOYAIANAHLATEKDKES---LHK 117  
 QY 118 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 Db 118 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 QY 120 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 Db 120 LLLQGLDITLIGSNDFTSYRNOIESAGLPLTPRESLGTLPAPASTTFNNAESNGA----NS 173  
 QY 174 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 Db 174 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 QY 180 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 Db 180 KPEVAKTGLGSSAAMTAVVAALHLYGVNDSLDPCKEGRGCS--DLDVTHMIAQTSHC 231  
 QY 232 LAQKVGSGDFVSCAVYGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 291  
 Db 232 LAQKVGSGDFVSCAVYGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 291  
 QY 240 IAGKIGSGDFVSAAYVGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 297  
 Db 240 IAGKIGSGDFVSAAYVGSQRYVRFSPPEVLSFAQVAVTGLPLNEVIGTILKGMKDNKTE 297  
 QY 292 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 Db 292 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 QY 298 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 Db 298 FSLPPLNMLFLGEGSGSTSPSWGAVKKWQMSDPEKARWONLS DANLETKLNDL 351  
 QY 352 SKLAKDHMDVLYRIKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILMONG 411  
 Db 352 SKLAKDHMDVLYRIKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILMONG 411  
 QY 358 KGLSENHREAVESVYLSCHLTYEKTWEVATNOHQLIKSLAARDACLETIRLHMREMG 417  
 Db 358 KGLSENHREAVESVYLSCHLTYEKTWEVATNOHQLIKSLAARDACLETIRLHMREMG 417  
 QY 412 EASVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 Db 412 EASVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 QY 418 IAGVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 Db 418 IAGVPIEPESQRLDITMNDGVLLAGVPGAGFDVAFVATLGDSTGNVAKWSSLNALLVRED 478  
 QY 472 LALLVREDPHVCLESQDPRTCITSGVSSIHL 504  
 Db 472 LALLVREDPHVCLESQDPRTCITSGVSSIHL 504  
 QY 478 LPLLVREDPHVCLESQDPRTCITSGVSSIHL 510  
 Db 478 LPLLVREDPHVCLESQDPRTCITSGVSSIHL 510

# RESULT 6 US-09-708-427-83408

; Sequence 83408, Application US/09708427  
 ; GENERAL INFORMATION:  
 ; APPLICANT: N. ALEXANDROV et al.  
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
 ; FILE REFERENCE: 2750-1243P  
 ; CURRENT APPLICATION NUMBER: US/09/708,427  
 ; CURRENT FILING DATE: 2000-11-09

```

? NUMBER OF SEQ ID NOS: 85364
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 83408
? LENGTH: 499
? TYPE: PRT
? ORGANISM: Zea mays subsp. mays
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1..499
? OTHER INFORMATION: Xaa is any amino acid
? NAME/KEY: misc_feature
? LOCATION: 1..499
? OTHER INFORMATION: Ceres Seq. ID 1962605
US-09-708-427-83408

Query Match          59.3%; Score 1531; DB 21; Length 499;
Best Local Similarity 61.0%; Pred No.3.2e-142;
Matches 313; Conservedity 68; Mismatches 108; Indels 24; Gaps 6;

OY 1 MAVVASAPGKVLMTGGYLVLEKPNAGLVISTNARFYAIKPINEEYKPPSMAKMTDVKL 60
Db 1 MEVVASAPGKVLIAAGYLVLEKPNAGLVISTNARFYAVRPLDLSLPADAMAMAMTDVYK 60
OY 61 TSPQISREBEMKLSLNHLTLOSASDSKNPVEHAIQYALAAHLATEKDKES---LHK 117
Db 61 TSPQISREBEMKLSLNHLTLOSASDSKNPVEHAIQYALAAHLATEKDKES---LHK 117
OY 118 LLLGGIDITILGSPNFYSYRNQIESAGLPPLPESLGTLPAPFASITFPAESNGA----NS 173
Db 120 LLLGGIDITILGSPNFYSYRNQIESAGLPPLPESLGTLPAPFASITFPAESNGA----NS 173
OY 120 LLLGGIDITILGSPNFYSYRNQIESAGLPPLPESLGTLPAPFASITFPAESNGA----NS 173
Db 120 LLLGGIDITILGSPNFYSYRNQIESAGLPPLPESLGTLPAPFASITFPAESNGA----NS 173
OY 174 KPEVAKTGLGSSAAMTTAVVAALLHYLGVVLDSDPCKEKGFGCS--DLDIYHIAQTSHC 231
Db 180 KPEVAKTGLGSSAAMTTAVVAALLHYLGVVLDSDPCKEKGFGCS--DLDIYHIAQTSHC 231
OY 232 LAOGVGVSEFDVSCAVGSGORVYRSPSEVLSPFAQVAVTGLPLNEVIGTILKGWMDKRT 291
Db 240 LAOGVGVSEFDVSCAVGSGORVYRSPSEVLSPFAQVAVTGLPLNEVIGTILKGWMDKRT 291
OY 292 FSLPPLMLNLFLGEPGSGSSTPSPMGAVKWKQMSDEKARENQNLSDANLELETKLNDL 351
Db 298 FSLPPLMLNLFLGEPGSGSSTPSPMGAVKWKQMSDEKARENQNLSDANLELETKLNDL 351
OY 352 SKLADHDMDVYLRYKSCSVLTSEKVVHLATEPINALIITKELLEAEAMLRIRIIRONG 411
Db 358 SKLADHDMDVYLRYKSCSVLTSEKVVHLATEPINALIITKELLEAEAMLRIRIIRONG 411
OY 412 EAAVYPIEPESQTOILLDSMTSAGVYLLACVPGAGGEAIFATILTGSGTKLITQAMSSHHV 471
Db 406 IAAVYPIEPESQTOILLDSMTSAGVYLLACVPGAGGEAIFATILTGSGTKLITQAMSSHHV 471
OY 472 LALLVREDPHVCYLCESGDPRTTCITSGVSSIHL 504
Db 466 LPLVREDPHVCYLCESGDPRTTCITSGVSSIHL 498

RESULT 7
US-09-708-427-83407
? Sequence 83407, Application US/09708427
? GENERAL INFORMATION:
? APPLICANT: N. ALEXANDROV et al.
? TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
? FILE REFERENCE: 2750-1243P
? CURRENT APPLICATION NUMBER: US/09/708,427
? CURRENT FILING DATE: 2000-11-09
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 83407
? LENGTH: 535
? TYPE: PRT
? ORGANISM: Zea mays subsp. mays
? FEATURE:

```

```

; NAME/KEY: misc-feature
; LOCATION: 1..535
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc-feature
; LOCATION: 1..535
; OTHER INFORMATION: Ceres Seq. ID 1962604
US-09-708-427-83407

Query Match          59.3%; Score 1531; DB 21; Length 535;
Best Local Similarity 61.0%; Pred. No. 3,7e-142;
Matches 313; Conservative 68; Mismatches 108; Indels 24; Gaps 6;

QY      1 MAVVASAGKYLMTGGYLLEKPNAGLVLTINARFYALVKPDINEEVKPESSAMKWTDVKL 60
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB      37 MEVVASAGKVLIGAGYLVLERPNAGLVTSTAFEFAYVRPLRDSLPADAAWMAFMDVKV 96
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY      61 TSPQLSRSMWKLSLNHLTLOSASASDSRNPFVEHAIOYALAIAAHLATKKDKES---LHK 117
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB      97 TSPQLSRRTATKLLSTKGALDQLTSARESTNPFVQAOLQFSVAAK--ATTIDKEKDYYDK 155
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     118 LLLOGIDITTIIGSNDIFYSYRNQIESAGLPTPEISGTLAPASTIFPNAAESNGA-----NS 173
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     156 LLLOGLNTTTIIGCNDIFYRKQIERGRPLTPPALSLIPRSSITTFPSEVANGMTEBKC 215
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     174 KPEVAKTGSGSAAMTTAVVAALLHYLGVDLSDPCKEKGKGC--DLDTVMHTAQTSHC 231
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     216 KPEVAKTGLGSSAAMTTSVVAALLHYLGVALSCPGSGSGDNTTRREIDLTHYSIAQSANC 275
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     232 LAAGVGSGFGPVSCAVVSORSRYVRPSPEVLFAQVAVTGLPLNTLVIGILTKGDKNK RTE 291
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     276 IAQGTVGGFGFDNAVNVSQRKRVRFSPFLSSQATGTGTFDP--DVYSDIVQARDHMKQ 333
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     292 FSLPLPLMMLFGEDESGGSSTPSWGA VKKMQMSDPEKARENMONLSDAELETKRLNDL 351
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     334 FSLPLPLMTLLIGEPCGTGSSSTPSWMSGSVKRMQKSDPECKPKTWSKLAFANALSLENQRL 393
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     352 SKLAADHDVLYLRVYIKSSSVLTSEKVYLHAREPINEAIKILLAREAMLPIRLMRONG 411
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     394 KGLSENHREAV-----ESWTEVATNQHOELTIKSLKLARDCLETIRLHMREM 441
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     412 EAASVIPPEISOQTOLDSTMSAEGLLAGVAGAGGFDAIPFTILDGSGTKLTLCAMSSHNV 471
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     442 IAAQVPRIPEDSQTRLDATNMMEGYLLAGVPGAGGFDAVFSVTLGDASNAAVANMSSAGV 501
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

QY     472 LALVREDPHGVCLSEGPRTTCITSGVSSIHL 504
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |
DB     502 LPLVIREDCRGVSELDADPRTREWSAAVSSIOT 534
        | | | | | | | | | | | | | | | | | | | | : : : | | | | |

RESULT 8
US-09-708-427-66289
; Sequence 66289, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66289
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..535
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc-feature
; LOCATION: 1..535
; OTHER INFORMATION: Ceres Seq. ID 1930323
US-09-708-427-66289
```

Query Match	52.0%;	Score 1341;	DB 21;	Length 535;
Best Local Similarity	59.6%;	Pred. No. 2.9e-123;		
Matches 277; Conservative	61;	Mismatches 103;	Indels 24;	Gaps 6

QY	49	ESMAKMTDVLTSPPOLSRBSMYL:LNHLTLOSVSASDRNPFVEHAIOYATAAHLAT	108
Db	85	DAKMAKMTDVKVYTPOLSRBETTKYLSLTKSALDQJLTSRBESTNPFVEADIDVYAAAK-KAT	143
QY	109	EKDKES---LHKLLLOGLDITITLGSNDPFSYRNOIESAGPLPTPEISGTLAPASITFNA	1655
Db	144	IIDREKDVYDXLLXGINTITLGCNDPFSYRQIEARGPLPLPEALLSLPPRSTTFNS	2033
QY	166	AESGCA----NSKEVAKTGLGSSAAATTAVVAALLHYLGVDLSDPCKGKGCSS--DL	219
Db	204	EVANGTMTGCKREKVAKTGIGSSAAATTSVAAALLHYLGVSILSCPGSSGDNTRREL	2633
QY	220	DVIMIAQTSHCLQGVKSGFDFVSCAVYGSORVYRFRSEPVLSFAQVAVGLPLNVEYGT	279
Db	264	DLYHSIQSANCINQKIGSGFDVSAAYVGSQRKYRSPPELLSSQARGGFLP--DYVSD	321
QY	280	ILKCKMDKRTESLPPLMLNLFLGEPGSGSSPTFSYGAVKWKQMSDPEKARENMONLSD	339
Db	322	IYTRQMHENKQFSLPLMLTLLGEPGTGSSPTFSYMGYSVKRMQKSDPECKDPTWSKLAF	381
QY	340	ANLEPTRLNDLSLADMDHVDYLARVAKSSVLTSEKVLHATEPINEALITIKELTEREA	399
Db	382	ANSLALENDLRLKLGISEKHREPAY-----ESWTEVATNQHOELLITKSLLAARDA	429
QY	400	MLRITLIROMGEAASVPIEPESQOTLLDSTMAEGVILLAGVGAAGFDAIFAITLGDSC	459
Db	430	CLEIRLHMRREGIAGVAPIEPDSQTRLLDATMNMEGVLLAGVGAAGFDAVSVTLGDAS	489
QY	460	TKLLOAMSSHHVALLVYREDPHGCLESQDPRTTCITISYVSSIHL	504
Db	490	NAVANAAMSSAGVLPPLVREDCRGVSLDADPRTREVSAAVSSIOI	534

```

RESULT 9
US-09-708-427-66290
: Sequence 66290, Application US/09708427
:
: GENERAL INFORMATION:
: APPLICANT: N. ALEXANDROV et al.
: TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
: FILE OF INVENTION: THERREBY
: FILE REFERENCE: 2750-1243P
: CURRENT APPLICATION NUMBER: US/09/708,427
: CURRENT FILING DATE: 2000-11-09
: NUMBER OF SEQ ID NOS: 85364
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 66290
: LENGTH: 326
: TYPE: PRT
: ORGANISM: Zea mays subsp. mays
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..326
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: misc_feature
: LOCATION: 1..326
: OTHER INFORMATION: Ceres Seq. ID 1930324
: US-09-708-427-66290

```

```

Query Match          37.7%  Score 973  DB 21  Length 326
Best Local Similarity 58.8%  Pred. No.5,1e-87
Matches 198; Conservative 45; Mismatches 78; Indels 16; Gaps 3
QY .170 GANSKPEVAKTGLGSSAMTTAVVALLHYLGVDLSPDCKEKGKFGCS--DLDVIHMIQ 227
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  3  GEKCKPEVAKTGLGSSAMTTSVVALLHYLGVDLSPDCKEKGKFGCSGNTTRREDIDVHSIQ 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 228 TSHCLAQGVSGFVSCAVVIGSQRRVRRSPVLSFAOVAVATGLPINEVYIGTILKGGKWD 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      63  SAHCIAQCKISSGDFVSAANYGSGRYRYRFPSEILLISQATGGTFLP--DVVSDIVTQWMDH 120
Oy      288  KTEPSELPPLMNLFLGEPGSGSSPTSMGAVKWKQMSDPEKARENKQNSDANILETK 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121  ENKFOSEPLPMTLLGEPGTGSSPTSMGVSVKRWMQSDPECKCDTWSKLAFANSALENO 180
Oy      348  LNDISKLAKHDWVLYLRVYKSCSVLTSEKWLHATPELINEAIITKELEAREAMRIRILM 407
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      181  LRLKGLSENIREFX-----ESKTEVATNQHOLEIITKSLAANDACLIEIRLHM 228
Oy      408  RQMGFAASVPIEPESQOTLDDSTNSAECVLLAGVPGAGGFDALFAITLSDSGTKLTOAMS 467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      229  REMGIAGVPIEDPSQOTRLDATNMKGCVLLAGVPGAGGFDAAFSVTLGDASNAVANAMS 288
Oy      468  SHNVATILIVREDPHGVCLSCSDPPTCTOISVSYIHL 504
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      289  SAGVLPFLIVREDCKVSLDEDADPTREVSAAVSSIQI 325

```

```

RESULT 10
US-09-708-427-83409
: Sequence 83409, Application US/09708427
: GENERAL INFORMATION:
: APPLICANT: N. ALEXANDROV et al.
: TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
: TITLE OF INVENTION: THEREBY
: FILE REFERENCE: 2750-1243P
: CURRENT APPLICATION NUMBER: US/09/708,427
: CURRENT FILING DATE: 2000-11-09
: NUMBER OF SEQ ID NOS: 85364
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 83409
: LENGTH: 326
: TYPE: prt
: ORGANISM: Zea mays subsp. mays
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..326
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: misc_feature
: LOCATION: 1..326
: OTHER INFORMATION: Ceres Seq. ID 1962606
US-09-708-427-83409

```

[illegible]

Query Match	34.9%	Score 902:	DB 21:	Length 306;
Best Local Similarity	57.4%	Pred. No. 5.2e-80;		
Matches 183;	Conservative 45;	Mismatches 75;	Indels 16;	Gaps 3

RESULT 12  
US-10-219-999-55015

1 GENERATED INFORMATION  
2 APPLICANT: Cao, Yongwei  
3 APPLICANT: Edgerton, Michael D  
4 APPLICANT: Hinkle, Gregory J.  
5 APPLICANT: Kovalic, David K.  
6 APPLICANT: Liu, Jindong  
7 APPLICANT: Stein, Joshua  
8 TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
9 FILE REFERENCE: 38-10/527267C  
10 CURRENT APPLICATION NUMBER: US/10/219,999  
11 CURRENT FILING DATE: 2002-08-15  
12 PRIOR APPLICATION NUMBER: US 60/324,109  
13 PRIOR FILING DATE: 2001-09-21  
14 PRIOR APPLICATION NUMBER: US 60/312,544  
15 PRIOR FILING DATE: 2001-08-15  
16 NUMBER OF SEQ ID NOS: 63520

Query Match	33.6%	Score 868.5	DB 26	Length 286
Best Local Similarity	59.2%	Pred. No. 1e-76		
Matches 170	Conservative 45	Mismatches 71	Indels 1	Gaps 1

Db 240 ASNAVAHAWSSAGVLP LLVREDCRGVSL EDADPTREVSAAVWSIQI 286

Query Match	32.5%	Score	839.5	DB	26	Length	281
Best Local Similarity	59.3%	Pred	NO. 7.5e-74				
Best 166, Conservative	43	Mismatches	70	Indels	1	Gaps	1

Oy	225	IAOISHCLAQKVGSGFVSAAYVGSQRXYAFSEDEVLTSFAOVAATGRLYMEIOTITLKGK	284
Dd	2	IAOSHACLAQKIGSGFVSAAVYGSOXRYAFSEPELSSAO - AIGGVLPDVPVSDVLTOR	60
Oy	285	WDNKTEPFLPLMLFLGEPGGSGSSPTPVWGAVKKKOMSDPEKARENNONLSDAIMEL	344
Dd	61	WDHEKQKSTLEPLMTLLGLGEPTGGSSPTSPMVGSYKRKLKSDPEKSDTMSLIAINSTL	120
Oy	345	ETKLNDLSKLAKHDVDYLRYIKSCSVLTSEKWVLHATEPINEAIIKELLEAREAMLIR	404
Dd	121	ENGRILKLGSEHNHEAVESWVRSCSRTLTYCKMAEVATANHOEOLITSLAARDACLEIR	180
Oy	405	IIMRMGEASVAPLEPESQTOLDISTMSAEVLLAGYPGAGGFPAITAITIGDSGTLYTO	464



Db 181 LHMREMGIAAGVIEPDSQTRLLDATNMNEGVLGAGVAGFDVAVSLGADSNAAVAH 240  
 QY 465 AMSSHNVALLVREDPHGVCLSGDPRCTCITGVSSTIHL 504  
 Db 241 AMSSVGVLLPLVREDCRGVSLDADPRTRVSAVWSIQI 280

# RESULT 14 US-60-312-544-6223

; Sequence 6223, Application US/60312544  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Edgerton, Michael D  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Stein, Joshua  
 ; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
 ; FILE REFERENCE: 38-10(52726)A  
 ; CURRENT APPLICATION NUMBER: US/60/312,544  
 ; CURRENT FILING DATE: 2001-08-15  
 ; NUMBER OF SEQ ID NOS: 10730  
 ; SEQ ID NO 6223  
 ; LENGTH: 281  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 700571779\_FLI  
 US-60-312-544-6223

Query Match 32.5%; Score 839.5; DB 27; Length 281;  
 Best Local Similarity 59.3%; Pred. No. 7.5e-74;  
 Matches 166; Conservative 43; Mismatches 70; Indels 1; Gaps 1;

QY 225 IAQSHCLAQKVGSGFDVSCAVYGSQRYVRESPEVLSFAQVAVTGILPNEVIGTILK 284  
 Db 2 IAQSAHCLAQKIGSGFDVSAVYGSQRYVRESPELSSAQ-AIGGTVLPDVVSDVLTQR 60  
 QY 285 WDKKREFSLPRLMNLFLGPGSGSSPSMGAVKWKQMSDPEKARENNONSDALEL 344  
 Db 61 WDHNKQFSLPRLMTLLGPGSGSSPSMGVSKWMLKSDPEKSRDTWSKLAIANSTL 120  
 QY 345 ETKLNDLSKLAKDMVDYLIVKSCSVLTSEKWLHATEPINEALIKELREAREAMLRI 404  
 Db 121 ENGLRLIKGLSENHHEYSMAVSCSRITGKMAEVAATNOHOLIRSLAARDACLEIR 180  
 QY 405 ILMRQGEAASVPIEPESQTLDDSTMSAGVLLAGVAGFDPAIFAILGDSGTQLQ 464  
 Db 181 LHMREMGIAAGVIEPDSQTRLLDATNMNEGVLGAGVAGFDVAVSLGADSNAAVAH 240  
 QY 465 AMSSHNVALLVREDPHGVCLSGDPRCTCITGVSSTIHL 504  
 Db 241 AMSSVGVLLPLVREDCRGVSLDADPRTRVSAVWSIQI 280

## RESULT 15

US-09-708-427-65887  
 ; Sequence 65887, Application US/09708427  
 ; GENERAL INFORMATION:  
 ; APPLICANT: N. ALEXANDROV et al.  
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
 ; FILE REFERENCE: 2750-1243P  
 ; CURRENT APPLICATION NUMBER: US/09/708,427  
 ; CURRENT FILING DATE: 2000-11-09  
 ; NUMBER OF SEQ ID NOS: 85364  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 65887  
 ; LENGTH: 359  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays subsp. mays  
 ; FEATURE:

; NAME/KEY: misc:feature  
 ; LOCATION: 1..359  
 ; OTHER INFORMATION: Xaa is any amino acid  
 ; NAME/KEY: misc:feature  
 ; LOCATION: 1..359  
 ; OTHER INFORMATION: Ceres Seq. ID 1929648  
 US-09-708-427-65887

Query Match 28.0%; Score 722.5; DB 21; Length 359;  
 Best Local Similarity 66.2%; Pred. No. 5e-62;  
 Matches 133; Conservative 26; Mismatches 41; Indels 11; Gaps 5;

QY 49 ESWAMKWTDVKLTPQSLRESMYKLSLNLTLQSVASDSRNPVEHAIOYAIANAHLAT 108  
 Db 114 DSWTMAWTDVKXTSPQSLRVATYLSLNTKTLQTLTSRESNPNPEQAIQPSVAAAK-AT 172  
 QY 109 EKDKES---LHKLLOGDITILGASNPFYSRNOIESAGLPITPESIGTIAPFASITFNA 165  
 Db 173 IIDKRDVVDKLLDGLNITILIGHNDFYSRKQIEARGLPITPEVLLSLPPESSITFNS 232  
 QY 166 AESNGA---NSKPEVAKTGIGSSAAMTTAVVALLHYLGVDLSDPCDKGKFGCS--DL 219  
 Db 233 EVANGTMTGEKCKPEVAKTGLGSSAAMTTSVYVALLHYLGAVNLSCFGQSSGDNAGREL 292  
 QY 220 DVYHMAQTSCLAQKVGSGFDVSCAVYGSQRYVRESPEVLSFAQVAVTG 270  
 Db 293 DIVHTINQSAHCLAQKIGSGFDVSAVYGSQRYVRESPELSSAQ-AIGG 342

Search completed: April 26, 2003, 13:00:27  
 Job time: 151 secs

THIS PAGE BLANK (user)





```

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3740
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3740

Query Match          20.5%; Score 529; DB 6; Length 402;
Best Local Similarity 33.8%; Pred. No. 1,6e-39;
Matches 172; Conservative 59; Mismatches 134; Indels 144; Gaps 23;

OY      6 SAPGKVLMTGGVYLEKPNAGLVLTARNARYAIKYPINEEKKPESMMAMKWTDVKLTSPDL65
DB      3 SAPGKVLLAGGYLVDNRNTGLVFGLSARLNIVSRDIQ-----ASPGVINHEIIVRSKOF 57
        ||||||| :|||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      66 SRESM---YKLS-----LNHLTLQ-SVSASDRNPFEVAHIOYAIA-----AAHLATEKD 111
DB      58 LKAEMRYGHLEADEDGEIKITQLQGGSASAKGNPFVETTLNVALTVTRRAAQTTO-- 115
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      112 KESLHKLLQGDITILGSDNFYSRNQIESAGLPILPEISGITLAFFASTTFMAAESGA 171
DB      116 -----TLKPVTLTLLADNDYTS-----SPTN-----NSNMAAGKGS 146
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      172 NSK-----PEVKTGTIGSSSAAMTTAVVAALL-HYL-GVYDLDSPCKEGKFGCSDL 219
DB      147 GGSRFAAYSTLEDARKTGTGGSSALVTALTALSLSHYLPDSLTDLA--TDEGK----- 198
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      220 DYIHMAOTSHCLAQSGFSCAVYGSQRVRSPEYLSFAOVAVTGLP-----L 273
DB      199 RLHHNLQAQAHCAOGKVGSGFDVAAAAYGSSHRRSPSLTS--SLPEACKPGFSAKLF 256
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      274 NEVIGTILGKMWNK--RTEFSLPLMLNFLGEPGSGSSPTPSWGAVKMKOMSDPEKAR 331
DB      257 SYVNGCDASSQMDTEYVKDAVSLPKGVAAKMCYDC-GSQTVGMYKVOLANRAAPKEAK 315
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      332 EWMQNLS DANLELETKLNDLSKLAKDHWDVLYRIKSCSVLTSEKWLHATEPINEALIK 391
DB      316 ELMDLEIQRRN-----EKLA-----SYLKEGK-----TE----- 338
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      392 ELLEAREAMLRIIRLMRWGEASVPTEPESOTOLDL-TMSAGCVLLAGVPGAGGPAT 450
DB      339 ---ELRPAAHAIRELVKRKGTEGSGVPLEPDSOKELLDALLEGVEGYGVGAGGYDA- 394
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      451 FAITLGDSGTKLQAWSSHNVALLVRED 479
DB      395 -----LALLVRDD 402
        : : : : : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-10-369-493-2162
; Sequence 2162, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

```

```

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2162
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2162

Query Match      18.8%; Score 485; DB 6; Length 423;
Best Local Similarity 30.1%, Pred. 0.1,8e-35;
Matches 151; Conservative 83; Mismatches 167; Indels 100; Gaps 21.

OY 6 SAPRGVMTGGGLGVLEKRPAGLVLSTNAFAYLVRPINEEVPESWAMKMDVKLTSPQ 64
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 3 SAPRGVLLAGVIVLDPOYSGLVIGLTKAGVASTTTD-----KCGIVRKSPGF 53
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 65 LSRESMYKLISLN-----HLTQSYVASDSNRPEVEHAIQVIAAANHLATEKDKS LKL 118
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 54 INAEMLYNIDMTVSPDIRVHQIYENCELEKNRPFQLALFYVI-NYFESTGR-----DPL 107
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 119 LLGGADITILGSDNYSVYRNQIESAGLPRLTPESLGTLAP---FASTFNAAESNGANSKP 175
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 108 CMQDQVLTLYQVDNAY-HQPQ-----LKPRQ--TSYKRFNFCTLG----- 146
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 176 EYAKTGSSGAAMTAAYVAALLHYLVGDVLSDPG--KEKGFCSCDLVDVTIMIAQSHCIA 233
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 147 GVHKHGLGSSAMITSLISTL--FLSRLRLTDGTGDKSLKIDSIRKIVHNLAQIAHCSA 204
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 224 QSKVGSGPVGSAAVYGQQRYRESEPVLSFAOV---AVTGLRLNEVICTLLGKWMDNR 289
    ||||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 205 QSKVSSGEFVGATVWGSCIYRRFDKRLLEQLLPPDEQIKNTFSTEKKIVSKMWDV- 263
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 290 TEFSPLPLNLFLFGPGSGGSTPSNVGAVKMKWMQSDPEKARENOMNSDANILETKLN 349
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 264 VPFQLPATYCILMGDY-AGGSSTPGWKVKVQOKQENPEEK-----NCRD----- 308
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 350 DLSKLAKHDWVLYKY--IKSCSVLTSEKMWVLAHEPIINELAIKELLEKEMLRIRILA 407
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 309 -----DLSRYLSLTKNCFL-----SSSIDSELQSOFRSIRIRLQRITY-- 347
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 408 ROMGEASVPIPEEQOTOLDSTMSAEGVLAGVPGAGGFDAIFATLGDG--TRLTQA 465
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 348 -----EAKVDIEPLKQTNRLNDNIEDLPQYIGVGVGAGGFDAQFCLATNHTEIENVKT 402
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
OY 466 WSSHVLLALLVREDPHGCYLE 486
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 403 WKDDGVPMVDVSPAFDGLAVE 423

RESULT 6
US-09-675-784A-13279
; Sequence 13279, Application US/09675784A
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLTING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: ASPERGILLUS FUMIGATUS NUCLEIC ACIDS AND POLYPEPTIDES,
FILE REFERENCE: 2976-4020US1
CURRENT APPLICATION NUMBER: US/09/675,784A
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/156,338
PRIOR FILING DATE: 1999-09-29
NUMBER OF SEQ ID NOS: 13925
SEQ ID NO 13279
LENGTH: 491
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-09-675-784A-13279
```

Query Match	17.6%	Score 453.5	DB 5	Length 451
Best Local Similarity	30.1%	Pred. No. 1.5e-32		
Matches 155	Conservative 72	Mismatches 185	Indels 103	Gaps 18
6	SAPGVLMTGCVLTVEKPNAGLVLSNNAFFVATVPRINE-----EVRPESSWAKK-----W	55		

```

Db 8 SAPKALLAGGYLVLTDRKEAFVYGLSARHAAVAHPYSGLSGDSKDEVRNKSQCFKDGEM 67
QY 56 TDVAKLTSPQLSRESMYKLSLNHLTLQS-----VSASDSRNFVFNHAIOYALAAHNLATEND 111
Db 68 -----LYHISPKSGFLIPVSIIGSKNPEIKVIANVFSTF----- 101
QY 112 KESLHKLTLGLDITLTILGSNDPFSYRNOISAGILPLTPESLITLAPFASITFNMAESNGA 171
Db 102 KPNNDYCNRLVYIDLFSD--AYHSQDS-----VTEHRG-----NRRLSFHS 144
QY 172 NSKPEVAKTGLSGSSAAMTTAVVAALLHYLGLVLDLSDPCKEGKFGCSLDLVIHMTAOTSIC 231
Db 145 HRIEVPKRTGIGSSAGLVLTALASFF--VSDLENNVDKXR-----EVIHMLQAVHC 197
QY 232 LAQGVSGEDVSCAVYGSORIVRFSEVLSFPAQAVTGLPLNEVIGTILKGR----- 284
Db 198 QAOCKISGDFVAAAAAGSIRYRFRPALIS-----NLP---DIGSATYSGKLAAHLD 247
QY 285 ---NDKNRTESLPLNNLFLGEPGSGSSSTPSNVGAVKKQMSDPEKARENONTSDAN 341
Db 248 EEDNNITTKSNHLPSSGLTMMGDI--KNGSETVKLVQKVKWYSHMPEISKITTELIDHAN 306
QY 342 LELETKLNDLSKLAKDMDVYLRYIKSC--SVLTSEKVMVHATEPINALIINKELLEAREA 399
Db 307 SRPMDGLSKLDRLEHTHDYSDQIFESLENNDCCTCKY-----PITEVROA 353
QY 400 MLRILIRLROMGEASVPIEPESOTOLLDSSTMSAEVYLLAGVPGAGFDAIFAITLGDG 459
Db 354 VATIRFSFRKITKESGADIEPPVQTSLLDCQGLKGLVLTCLIPAGGDAIAVITRKQVD 413
QY 460 TKLTQAMSSH--NVLALLVREDPHGVCLESGDPT 492
Db 414 LRAQTANDKRFESKVQWLDVTOADMGVKRKE--DPET 447

RESULT 8
US-10-369-493-1909
/ Sequence 1909, Application US/10369493
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 1909
/ LENGTH: 451
/ TYPE: PRT
/ ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1909

Query Match 17.6%; Score 453.5; Db 6; Length 451;
Best Local Similarity 30.1%; Pred. No. 1,5e-32;
Matches 155; Conservative 72; Mismatches 185; Indels 103; Gaps 18;

QY 6 SAPKGLVMTGGYLVLEKPRNAGLVLTNARFYAIYKPKNE-----EYKPESNAMK---W 55
Db 8 SAPKALLAGGYLVLTDRKEAFVYGLSARHAAVAHPYSGLSGDSKDEVRNKSQCFKDGEM 67
QY 56 TDVAKLTSPQLSRESMYKLSLNHLTLQS-----VSASDSRNFVFNHAIOYALAAHNLATEND 111
Db 68 -----LYHISPKSGFLIPVSIIGSKNPEIKVIANVFSTF----- 101
QY 112 KESLHKLTLGLDITLTILGSNDPFSYRNOISAGILPLTPESLITLAPFASITFNMAESNGA 171
Db 102 KPNNDYCNRLVYIDLFSD--AYHSQDS-----VTEHRG-----NRRLSFHS 144
QY 172 NSKPEVAKTGLSGSSAAMTTAVVAALLHYLGLVLDLSDPCKEGKFGCSLDLVIHMTAOTSIC 231
Db 145 HRIEVPKRTGIGSSAGLVLTALASFF--VSDLENNVDKXR-----EVIHMLQAVHC 197
QY 232 LAQGVSGEDVSCAVYGSORIVRFSEVLSFPAQAVTGLPLNEVIGTILKGR----- 284
Db 198 QAOCKISGDFVAAAAAGSIRYRFRPALIS-----NLP---DIGSATYSGKLAAHLD 247
QY 285 ---NDKNRTESLPLNNLFLGEPGSGSSSTPSNVGAVKKQMSDPEKARENONTSDAN 341
Db 248 EEDNNITTKSNHLPSSGLTMMGDI--KNGSETVKLVQKVKWYSHMPEISKITTELIDHAN 306
QY 342 LELETKLNDLSKLAKDMDVYLRYIKSC--SVLTSEKVMVHATEPINALIINKELLEAREA 399
Db 307 SRPMDGLSKLDRLEHTHDYSDQIFESLENNDCCTCKY-----PITEVROA 353
QY 400 MLRILIRLROMGEASVPIEPESOTOLLDSSTMSAEVYLLAGVPGAGFDAIFAITLGDG 459
Db 354 VATIRFSFRKITKESGADIEPPVQTSLLDCQGLKGLVLTCLIPAGGDAIAVITRKQVD 413
QY 460 TKLTQAMSSH--NVLALLVREDPHGVCLESGDPT 492
Db 414 LRAQTANDKRFESKVQWLDVTOADMGVKRKE--DPET 447

```

[illegible]

```

RESULT 9          US-09-745A-20
US-09-909-745A-20
; Sequence 20, Application US/09909745A
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: Squalene Synthesis Enzymes
; FILE REFERENCE: Bb112 US CIP
; CURRENT APPLICATION NUMBER: US/09/909, 745A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/433, 242
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/107, 241
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-909-745A-20

Query Match      8.4% Score 216; DB 5; Length 67;
Best Local Similarity 60.0%; Pred. No. 2.8e-12;
Matches 39; Conservative 12; Mismatches 14; Indels 0; Gaps 0.

QY   309 GSSTPMGAVKWKQMSPDKARENWQNIISDANLEFRLKLANDLSKLAKDHWDVYLRAVTKS 368
      |||||:::||:||||:| | | | | | | :| | | | | | | :| | | |
Db    1  GSSTPMGSGVKOMQSDPKSKETWSKLGIANSVLENLRMLNKLAEDHWEAYESVLPSS 60
      |||||:::||:||||:| | | | | | | :| | | | | | | :| | | |
QY   369 CSVL 373
      || | |
Db    61 CSRLT 65

RESULT 10
US-09-909-745A-16
; Sequence 16, Application US/09909745A
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: Squalene Synthesis Enzymes
; FILE REFERENCE: Bb112 US CIP
; CURRENT APPLICATION NUMBER: US/09/909, 745A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/433, 242
; PRIOR FILING DATE: 1999-11-04

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	7.9%	Score 204	DB 5	Length 54
Best Local Similarity	74.1%	Pred. No. 2	4e-11	
Matches	40	Conservative	8	Mismatches 6
				Indels 0
				Gaps 0
397	RRAMRIRITLMOMGEAAVPTPEPSQQLDITSTASAEVILLIAGVAGGCPAI	450		
	:     :     :     :     :     :     :     :     :     :			
DB	1 RDACEIRIHRHREMGIIAGVPTPEPSQTRRLDITIMMEGVILLASVPGAGCPNAV	54		

```

RESULT 11
US-09-909-745A-18
: Sequence 18, Application US/09909745A
: GENERAL INFORMATION:
: APPLICANT: Falco, S. Carl
: TITLE OF INVENTION: Squalene Synthases Enzymes
: FILE REFERENCE: B8112 US CIP
: CURRENT APPLICATION NUMBER: US/09/909,745A
: CURRENT FILING DATE: 2001-07-20
: PRIOR APPLICATION NUMBER: 09/433,242
: PRIOR FILING DATE: 1999-11-04
: PRIOR APPLICATION NUMBER: 60/107,241
: PRIOR FILING DATE: 1998-11-05
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 18
: LENGTH: 54
: TYPE: prt
: ORGANISM: Zea mays
US-09-909-745A-18

```

```

Query Match Similarity      7.9%;   Score 204; DB 5; Length 54;
Best Local Similarity      74.1%;   Pred. No. 2,4e-11;
Matches      40; Conservative      8; Mismatches      6; Indels      0; Gaps      0;

OY      397  REAMRIRILRMQGEASVPIEPESQOTOLDSTSAEGVLLAGVPGAGCPAT 450
      1:1 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
Db      1  RDACLEIRLHMREMGIAAGVPIEPSQTRLLDATNMMEGVLLAGVPGAGCPADV 54

RESULT 12
US-10-282-122A--56929
: Sequence 56929, Application US/10282122A
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848

```

PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 56929  
LENGTH: 368  
TYPE: PR1  
ORGANISM: Enterococcus faecalis  
US-10-282-122A-56929

Query Match 6.9%; Score 177.5; DB 6; Length 368;  
Best Local Similarity 22.3%; Pred. No. 1.6e-07;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGLVSTNARFAIYKPINEEVKPSMAWKMTDVLTLS 62  
DB 2 IEVTPPKLFIAGYAYVEPCHPAIYAVDFVTVETDETSISAOYSSLP1RWT- 60  
QY 63 POLRESNYKLSLNHLTQSVASDSRNPFVEHAIOYAIAMHATEKDKESLKLLOG 122  
DB 61 ---RRNELVDIR-----ENPF-----HYLAIHL-TEKYAOEONKEL--- 96  
QY 123 LDITILSNDYFYSRNOIESAGLPLTPESLGTLPFASITPNAESNGANSKPEVAKTGL 182  
DB 97 -----SFYHLK-----VTSELDSSNG-----RKYGL 117  
QY 183 GSSAAMTTAVVAAL-LHYLGVDLSDPCKEKGCSLDVIMIAQTSCHLAQKVGSG 240  
DB 118 GSSGAVTVGTYKALNIFYDLG-----ENEIEFKLSALAHAVGNG-GSC 161  
QY 241 FDVSCAVYGSORRYRFSPEVLSPFAQVAVTGILPLNEVIGTILKGWDMKRT-EFSLPLMN 299  
DB 162 GDIAASCYG--WIAFS---TFPDHWVNOKVATETITDLAMQMPLEMTPLKVPKOLR 215  
QY 300 LFLGEPGSGSSTPSMVGAVKQWQSDPEKARENMQNLSDANLELETKLNDLSKLDHW 359  
DB 216 LLIGWTGS-PASTSDLVDRV--HOSKEKOAAEQFLMKSRCLCVETIMINGFN----- 264  
QY 360 DYLIRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILRMONGEAASVPIE 419  
DB 265 -----TKIS-VIOKQITKNQ-----LLAELSLTGVVIE 294  
QY 420 PESOTQLDSTMSAGVLLAGVPGAGFDALFATILGDSG-TKLTQAMSSHNVLAL 474  
DB 295 TEALKNLCDLAESYTG--AKSSGAGGDCGIVIFROKSGILPLMTAMENKIGITPL 348

RESULT 13  
US-09-134-000C-5356  
Sequence 5356, Application US/09134000C  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5356  
LENGTH: 370  
TYPE: PR1  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-5356

Query Match 6.9%; Score 177.5; DB 5; Length 370;  
Best Local Similarity 22.3%; Pred. No. 1.7e-07;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;

QY 3 VVASAPGKVLMTGGYLVLEKPNAGLVSTNARFAIYKPINEEVKPSMAWKMTDVLTLS 62  
DB 4 IEVTPPKLFIAGYAYVEPCHPAIYAVDFVTVETDETSISAOYSSLP1RWT- 62  
QY 63 POLRESNYKLSLNHLTQSVASDSRNPFVEHAIOYAIAMHATEKDKESLKLLOG 122  
DB 63 ---RRNELVDIR-----ENPF-----HYLAIHL-TEKYAOEONKEL--- 98  
QY 123 LDITILSNDYFYSRNOIESAGLPLTPESLGTLPFASITPNAESNGANSKPEVAKTGL 182  
DB 99 -----SFYHLK-----VTSELDSSNG-----RKYGL 119  
QY 183 GSSAAMTTAVVAAL-LHYLGVDLSDPCKEKGCSLDVIMIAQTSCHLAQKVGSG 240  
DB 120 GSSGAVTVGTYKALNIFYDLG-----ENEIEFKLSALAHAVGNG-GSC 163  
QY 241 FDVSCAVYGSORRYRFSPEVLSPFAQVAVTGILPLNEVIGTILKGWDMKRT-EFSLPLMN 299  
DB 164 GDIAASCYG--WIAFS---TFPDHWVNOKVATETITDLAMQMPLEMTPLKVPKOLR 217  
QY 300 LFLGEPGSGSSTPSMVGAVKQWQSDPEKARENMQNLSDANLELETKLNDLSKLDHW 359  
DB 218 LLIGWTGS-PASTSDLVDRV--HOSKEKOAAEQFLMKSRCLCVETIMINGFN----- 266  
QY 360 DYLIRVTKSCSVLTSEKWLHATEPINEAIIKELLEAREAMLRIRILRMONGEAASVPIE 419  
DB 267 -----TKIS-VIOKQITKNQ-----LLAELSLTGVVIE 296  
QY 420 PESOTQLDSTMSAGVLLAGVPGAGFDALFATILGDSG-TKLTQAMSSHNVLAL 474  
DB 297 TEALKNLCDLAESYTG--AKSSGAGGDCGIVIFROKSGILPLMTAMENKIGITPL 350

RESULT 14  
US-09-134-000C-5356  
Sequence 5356, Application US/09134000C  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5356  
LENGTH: 370  
TYPE: PR1  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-5356

Query Match 6.9%; Score 177.5; DB 5; Length 370;  
Best Local Similarity 22.3%; Pred. No. 1.7e-07;  
Matches 106; Conservative 58; Mismatches 179; Indels 133; Gaps 20;  
QY 3 VVASAPGKVLMTGGYLVLEKPNAGLVSTNARFAIYKPINEEVKPSMAWKMTDVLTLS 62  
DB 4 IEVTPPKLFIAGYAYVEPCHPAIYAVDFVTVETDETSISAOYSSLP1RWT- 62



```

Db 4 IEVTPGKLFAGEYAVVERGHPAIIIVADQFVTVEETDEGSISAOYSSLPJRM - 62
QY 63 POLSRESMYKLSLNHLTLQSVASDSNRNPFVEHAIOYAIAAHATEKRESLKLLQG 122
Db 63 ---RRNELVLDIR-----ENPF-----HYLAIRHL-TEKYAOEONKEL--- 98
QY 123 LDITILGSDNFYSYRNOIESAGLPLTPRESLGTLPFASITFNAESNGANSKPEVAKTGL 182
Db 99 -----SFYHLK-----VTELSDSNG-----RKGL 119
QY 183 GSSAAMTAAYVAAL--LHYLGVDLSDPCKEKGKFCSDLDVIMIAOTSICLAOGKYGSG 240
Db 120 GSSGAVTGVTAKLINFITYDGL-----ENERIFRLSALAHAYOGN-GSC 163
QY 241 FDVSCAVYGSORVYRFSPEVLSFAQAVVTGLPLNEVIGTILKGMKDKRT-EFSLPLPMN 299
Db 164 GDIAASCTGG--WIAFS---TFHDWVNOKVATEITLIDLAMPPELMFPLKVPKOLR 217
QY 300 LFLGEPGSGSSPSPNAGAVKQMSDPEKARENMQNLSDANLELETKLNDKSLAKDHW 359
Db 218 LLIGWTS-PASTSDLDVDR---HOSKEKQAEQOFLMSRLCVMINGFN----- 266
QY 360 DYLIRVTKSCSVLTSEKWLHATEPINEAIKELLEAREAMLRIIRLMONGEAASVPIE 419
Db 267 -----TGKIS-VIOKOITKNRO-----LLAELSLTGAVIE 296
QY 420 PESOTOLLDSMTSAGEVLAGVPAGGFDALFATLIGDSG-TKLQAMSSHNVAL 474
Db 297 TEALKNLCDLAESTYGA--AKSSGAGGDCGIVIFROKSGILPLMTAMEKDGITPL 350

```

## RESULT 15

```

PCT-US02-36122-42
; Sequence 42: Application PC/TUS0236122
; GENERAL INFORMATION:
; APPLICANT: Murphy, Ellen and Projan, Stephen, J.
; TITLE OF INVENTION: Allolococcus otitidis Infectious Disease Targets
; FILE REFERENCE: Application 1
; CURRENT APPLICATION NUMBER: PCT/US02/36122
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Allolococcus otitidis
PCT-US02-36122-42

```

Query Match 6.0%; Score 154; DB 1; Length 362;

Best local Similarity 20.8%; Pred. No. 2,2e-05;

Matches 103; Conservative 67; Mismatches 168; Indels 158; Gaps 20;

```

QY 8 PGKVLMTGGLVLEKPNAGLVLTNARFYAI---VKPINEVKP---ESMAKMTDV 58
Db 8 PGKLYLAGYAVVTPGYAGILL-TVSRVLTLDIWTETSPDQASVRSQTYGNOAYAMERLD- 65
QY 59 KLTSPQLSRESMYKLSLNHLTLQSVASDSNRNPFVEHAIOYAIAAHATEKDESLHKL 118
Db 66 -----GIFSRKDWHPF--HLVE----- 81
QY 119 LLAGLDTILGSDNFYSYRNOIESAGLPLTPRESLGTLPFASITFNAESNGANSKPEV- 177
Db 82 -----TVIQVTEAY-----TESLSLPL-----KSYGIGIQKSQLD 110
QY 178 ---AKTGLSSAAMTAVV--AALLHYLGVDLSDPCKEKGKFCSDLDVIMIAOTSHCL 232
Db 111 YQKKIGIGSSGAVTIIVIRGLSLYDLHLKLDID-----IFKLAIAIHQ 155
QY 233 ACGKVGSGFDVSCAVYGSORVYRFSPEVLSFAQAVVTGLPLNEVIGTILKGMKDKRT- 291
Db 156 LKSKGSGFDLACTYTGIRQSLDREMLQ-EQIS-----NHSIKDLAMPDPSLGIDR 208
QY 292 FSLPPLMLNLFGEPSGSGSSTPSMVGAVKKQMSDPEKARENMQNLSDANLELETKLNDL 351

```

```

Db 209 LSLPHDLRLIGWGTGQ--PASTEKLVQAVYP-----OKITRTPIDFQSFILDO- 253
QY 352 SKLAKDHDVYLIRVTKSCSVLTSEKWLHATEPINEAIKELLEAREAMLRIIRLMONG 411
Db 254 ---SOECVDGLVESLSQADSOASLAMI-----OKNRFLKAMG 288
QY 412 EAASVPIEPESOTOLLDSMTSAGEVLAGVPAGGFD-AIFATLIGDSGFKLTQAMSSHN 470
Db 289 QSRKQVITFKALTYLCLDIYAKYGG--QAKSSGAGGDCGIGLITRESPTEALYREMDAG 346
QY 471 VLAALLVREDPHGYCLE 486
Db 347 ILPLRLDIVENGACVD 362

```

Search completed: April 26, 2003, 13:01:55  
Job time : 84 secs

THIS PAGE BLANK (USPTO)